



Уральский
федеральный
университет

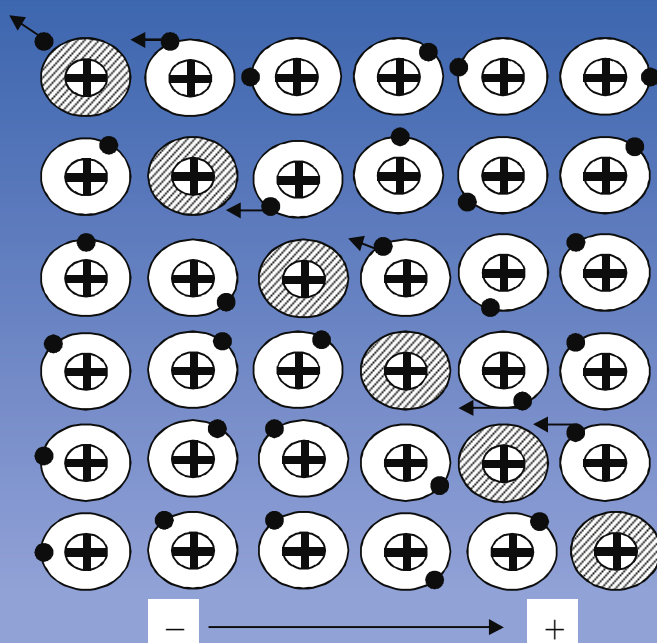
имени первого Президента
России Б.Н.Ельцина

Институт радиоэлектроники
и информационных
технологий — РТФ

А. А. ДУРНАКОВ

ФИЗИЧЕСКИЕ ОСНОВЫ МИКРО- И НАНОЭЛЕКТРОНИКИ

Учебное пособие



Министерство науки и высшего образования
Российской Федерации
Уральский федеральный университет
имени первого Президента России Б. Н. Ельцина

А. А. Дурнаков

ФИЗИЧЕСКИЕ ОСНОВЫ МИКРО- И НАНОЭЛЕКТРОНИКИ

Учебное пособие

Рекомендовано методическим советом
Уральского федерального университета
для студентов вуза, обучающихся
по направлениям подготовки
11.03.03 — Конструирование и технология
электронных средств, 11.03.01 — Радиотехника

Екатеринбург
Издательство Уральского университета
2020

УДК 621.382.049.77:53(075.8)

ББК 32.844я73

Д84

Рецензенты:

канд. техн. наук, доц., руководитель департамента дистанционного образования Уральского государственного колледжа им. И. И. Ползунова *В. В. Кийко*;

начальник отдела «НПО автоматики им. академика Н. А. Семихатова» *А. С. Наронов*

Научный редактор — проф., канд. техн. наук *А. А. Калмыков*

Дурнаков, А. А.

Д84 Физические основы микро- и нанoeлектроники : учебное пособие / А. А. Дурнаков ; Мин-во науки и высшего образования РФ. — Екатеринбург : Изд-во Урал. ун-та, 2020. — 247, [1] с.
ISBN 978-5-7996-2983-0

Пособие содержит описания физических эффектов и их компонентов, классификацию веществ, модели энергетических зон и ковалентных связей, физическое описание собственного и примесного полупроводника, применение и характеристики однородных полупроводников. Рассмотрено равновесное и неравновесное состояние p - n перехода, токи в нем, виды пробоев, туннелирование в сильнолегированных p - n переходах, гетеропереходы, вольтамперные характеристики переходов. В конце пособия приводятся практические и домашние задания.

Библиогр.: 13 назв. Табл. 4. Рис. 114. Прил. 1.

УДК 621.382.049.77:53(075.8)

ББК 32.844я73

ISBN 978-5-7996-2983-0

© Уральский федеральный
университет, 2020

ПРЕДИСЛОВИЕ

Использование полупроводников в электронике прошло длительный путь — от первого детектора на кристалле сернистого свинца до современных микропроцессоров. Такой результат достигнут благодаря успехам технологии, которая опирается на физическую электронику. В наши дни развитие микро- и наноэлектроники непрерывно стимулируется успехами в области физики полупроводников и в области технологии производства новых полупроводниковых структур.

Физической электроникой называют науку, которая занимается изучением и использованием потоков движущихся электронов, порождающих электрический ток, и так принято называть науку, изучающую электронные свойства некоторых твердых тел, кроме того, методы получения материалов с такими характеристиками, которые позволяют создавать устройства для передачи и накопления электронов. При этом рассматриваются не любые материалы, а лишь полупроводники, характеристики которых интересны с точки зрения технических приложений.

Дисциплина «Физические основы микро- и наноэлектроники» относится к группе естественно-научных дисциплин и ее целью является изучение физики электрических явлений в твердых телах. Особое внимание уделяется основам зонной теории твердых тел, физическим механизмам и математическому описанию основных (электрических, тепловых, оптических и магнитных) свойств равновесных, неравновесных полупроводников, особенностям контактов различных веществ, поверхностным состояниям твердых тел; рассматриваются различные физические эффекты, а также их применение в различных приборах и элементах. Пособие охватывает только часть материала, изучаемого в рамках дисциплины «Физические основы микро- и наноэлектроники».

1. ФИЗИЧЕСКИЙ ЭФФЕКТ И ЕГО КОМПОНЕНТЫ

1.1. ОПРЕДЕЛЕНИЕ ФИЗИЧЕСКОГО ЭФФЕКТА

Для однозначности толкования понятия «физический эффект» (ФЭ) принято следующее определение: физический эффект — это закономерность проявления результатов взаимодействия объектов материального мира, осуществляемого посредством физических полей. При этом закономерность проявления характеризуется последовательностью и повторяемостью при идентичности взаимодействия.

Все физические поля и их модификации будем рассматривать как воздействие в отрыве от тех материальных объектов, от которых они исходят.

Воздействие всегда направлено на некоторый материальный объект (в дальнейшем — объект), которым может быть отдельный элемент или совокупность взаимосвязанных элементов, образующих определенную структуру. Так, к объектам могут быть отнесены: системы из макротел (в т. ч. детали приборов, механизмов и др.), макротела (твердое тело, жидкость, кристалл и т. д.), молекула, атом, части атомов и молекул, частицы и т. д.

Результаты воздействия — это эффекты, проявляющиеся на объектах (или в окружающем их пространстве), на которые направлены определенные воздействия. Результатами воздействия являются те же физические поля, которые относятся к воздействиям. Этим обуславливается взаимосвязь между ФЭ, которая используется в объектах техники. К результатам воздействия относятся также измерения параметров объектов (размеров, формы, диэлектрической проницаемости и т. д.). При постоянстве условий взаимодействия и свойств объекта проявляются одни и те же результаты воздействия.

Схематическое изображение ФЭ (рис. 1) позволяет наглядно представлять физические процессы, происходящие при взаимодействии материальных объектов, в т. ч. в объектах техники.

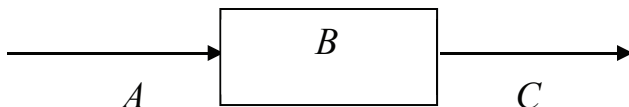


Рис. 1. Структурная схема физического эффекта;

A – воздействие; B – физический объект, на который оказывается воздействие;
 C – результат воздействия (эффект)

Анализ информации об известных ФЭ позволил выявить следующие закономерности их проявления.

- 1) При одном воздействии на объект может проявляться несколько результатов воздействия. Их число зависит от структуры физического объекта. Чем менее сложен по своей структуре объект, тем меньшее число результатов воздействия на нем проявляется. Изменяя число и состав структурных элементов, при заданном воздействии можно получить необходимые результаты воздействия.
- 2) На один объект может быть оказано несколько воздействий. Их можно подразделить на основные и дополнительные. Основным воздействием на объект является такое, результат от которого представляет другое, отличное от воздействия, физическое поле. Дополнительное воздействие — такое, которое приводит лишь к количественному изменению результата, получаемого от основного воздействия. Дополнительными воздействиями могут являться постоянно действующие воздействия, например гравитационное и тепловое поля, а также воздействия, оказываемые объектами окружающей среды. Так, на ФЭ возникновения магнитного поля при протекании электрического тока, где воздействием является электрическое поле, может быть дополнительно оказано воздействие, например, магнитным полем, что вызовет изменение электрического сопротивления проводника (магниторезистивный эффект). Дополнительные воздействия не могут вызвать данного результата без основного воздействия.
- 3) На одном объекте могут проявляться несколько ФЭ. Если какое-либо воздействие приводит к новому результату воздействия,

то налицо проявление на одном объекте двух разных ФЭ. Эта особенность является основой для многофункционального использования элементов объектов техники. Любые два ФЭ или более, проявляющиеся на одном объекте, оказывает влияние друг на друга. Это влияние объясняется участием структурных элементов объекта в проявлении обоих ФЭ.

- 4) Значение результата воздействия можно регулировать; оно обуславливается следующими факторами: количественным изменением воздействия, введением дополнительных воздействий, введением еще одного основного воздействия, изменением структуры объекта, изменением параметров объекта. Это свойство лежит в основе технической реализации таких функций объектов техники, как усиление, ослабление, преобразование, модуляция, выравнивание, проводимость, непроводимость и др.
- 5) При постоянстве условий воздействия и свойств объекта проявляются одни и те же результаты воздействия, одни и те же значения их параметров. Эта закономерность обусловлена стабильностью структуры объекта, на который оказывается воздействие, и стабильностью свойств элементов структуры.

1.2. ФИЗИЧЕСКИЕ ЭФФЕКТЫ В ЭЛЕКТРОНИКЕ

Физический эффект — это реальное явление, происходящее в физической системе и характеризующееся причинно-следственной связью между двумя или несколькими величинами. Уточним структуру физического эффекта.

В электронике чаще всего рассматриваются следующие воздействия: электрическое, механическое, магнитное, тепловое, оптическое, радиационное и акустическое.

Физические объекты электроники многообразны и могут быть классифицированы по ряду признаков, например:

- 1) по агрегатному состоянию — газ, жидкость, твердое тело;
- 2) по составу вещества — простые (германий, кремний, селен и др.) и составные (антимонид индия, арсенид галлия, карбид кремния и др.);

- 3) по структуре — однородные, неоднородные и контакты. Среди контактов можно выделить следующие — проводник-проводник, полупроводник-полупроводник (гомопереходы и гетеропереходы), металл-полупроводник, металл-диэлектрик-полупроводник;
- 4) по величине электропроводности — проводник, полупроводник, диэлектрик, сверхпроводник;
- 5) по величине магнитной проницаемости — ферромагнетики, парамагнетики, диамагнетики, ферримагнетики;
- 6) по виду вольтамперной характеристики — объекты с линейной, параболической, экспоненциальной, N -образной, S -образной характеристикой и т. д.

Результаты воздействия или эффекты, встречающиеся в электронике, также многообразны. Перечислим некоторые из них: изменения агрегатного состояния, электропроводности, силы и плотности тока, напряжения, магнитной индукции, магнитной проницаемости, температуры, давления; перемещение Δx , излучение, возникновение акустических волн, появление потока частиц, возникновение радиочастотных колебаний и т. д.

2. Однородный полупроводник

2.1. Собственный полупроводник

Классификация веществ

Все вещества по электрофизическим свойствам могут быть разделены на три больших класса: проводники, полупроводники и диэлектрики. Наиболее просто, казалось бы, классифицировать вещества по удельной электрической проводимости σ (или по удельному электрическому сопротивлению ρ). В создании электрического тока могут принимать участие только подвижные носители электрических зарядов. Поэтому электропроводность вещества тем больше, чем больше в единице объема этого вещества находится подвижных носителей электрических зарядов. В металлах практически все валентные электроны свободны, что и обуславливает их высокую электропроводность.

При $T = 300$ К у проводников удельная электрическая проводимость имеет значения 10^4 – 10^6 См/см (напомним, что 1 См/см есть проводимость 1 см³ вещества), у диэлектриков она меньше — 10^{-10} См/см, а у полупроводников ее значения находятся в пределах 10^{-10} – 10^4 См/см. Как видно, для полупроводников характерен очень широкий диапазон удельной проводимости. Поэтому значения удельной проводимости не могут служить в качестве однозначного критерия для классификации веществ.

Характерной особенностью полупроводников является ярко выраженная температурная зависимость удельной электрической проводимости. С повышением температуры, она, как правило, увеличивается на 5–6 % на градус, в то время как у металлов удельная электрическая

проводимость с повышением температуры падает на десятые доли процента на градус. Кроме того, удельная электрическая проводимость полупроводника также резко возрастает при введении в него незначительного количества примеси, а металлических сплавов — уменьшается.

Итак, полупроводник — это вещество, электропроводность которого занимает промежуточное положение между проводниками и диэлектриками; основным свойством этого вещества является сильная зависимость удельной проводимости от воздействия внешних факторов (температура, концентрация примесей, световое и ионизирующее излучение и др.). Большинство известных веществ относится именно к полупроводникам. К полупроводникам относятся элементы четвертой группы периодической таблицы Менделеева, а также химические соединения элементов третьей и пятой групп типа $A^{III} B^V$ (GaAs, InSb) и второй и шестой групп типа $A^{II} B^{VI}$ (CdS, BbS, CdFe). Ведущее место среди полупроводниковых материалов, используемых в полупроводниковой электронике, занимают кремний, германий и арсенид галлия GaAs.

Собственный полупроводник или полупроводник *i*-типа (от англ. *intrinsic* — собственный) — это идеально чистый полупроводник, т. е. влиянием примеси на его свойства можно пренебречь.

Модель ковалентной связи

Полупроводники, как правило, — твердые тела с регулярной кристаллической структурой — монокристаллы. Их кристаллическая решетка состоит из множества повторяющихся и примыкающих друг к другу элементарных ячеек той или иной формы и размера.

Собственные полупроводники имеют кристаллическую структуру, характеризующуюся периодическим расположением атомов в узлах пространственной кристаллической решетки. Взаимное притяжение атомов кристаллической решетки осуществляется за счет ковалентной связи, т. е. общей пары валентных электронов, вращающихся по одной орбите вокруг этих атомов. Согласно принципу Паули, общую орбиту могут иметь только два электрона с различными спинами, поэтому число ковалентных связей атома определяется его валентностью.

Рассмотрим модель ковалентных связей на примере германия или кремния (четвертая группа периодической таблицы Менделеева). Они образуют кубическую решетку типа алмаза, которая состоит из тетра-

эдров; расстояние между смежными атомами около 0,5 нм. В такой решетке каждый атом взаимно связан с четырьмя соседними атомами ковалентными связями (рис. 2), в результате влияния которых происходит обобществление валентных электронов и образование устойчивых электронных оболочек, состоящих из восьми электронов.

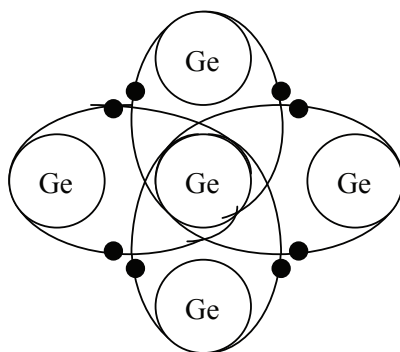


Рис. 2. Ковалентная связь между атомами германия

Как видно на рисунке, вокруг каждой пары атомов движутся по орбитам 2 валентных электрона, показанных жирными точками. В условном плоскостном изображении такой кристаллической решетки (рис. 3) ковалентные связи показаны в виде прямых линий, а электроны — по-прежнему в виде точек (иногда для упрощения электроны вообще не показывают).

При температуре абсолютного нуля ($T = 0^\circ \text{K}$) все валентные электроны находятся в ковалентных связях, следовательно, свободные носители заряда отсутствуют и полупроводник подобен диэлектрику.

При повышении температуры или при облучении полупроводника лучистой энергией, валентный электрон может выйти из ковалентной связи и стать свободным носителем электрического заряда. При этом ковалентная связь становится дефектной, в ней образуется свободное (вакантное) место, которое может занять один из валентных электронов соседней связи, в результате чего вакантное место переместится к другой паре атомов. Перемещение вакантного места внутри кристаллической решетки можно рассматривать как перемещение некоторого фиктивного (виртуального) положительного заряда, величина которого равна заряду электрона. Такой положительный заряд принято называть дыркой (рис. 4).

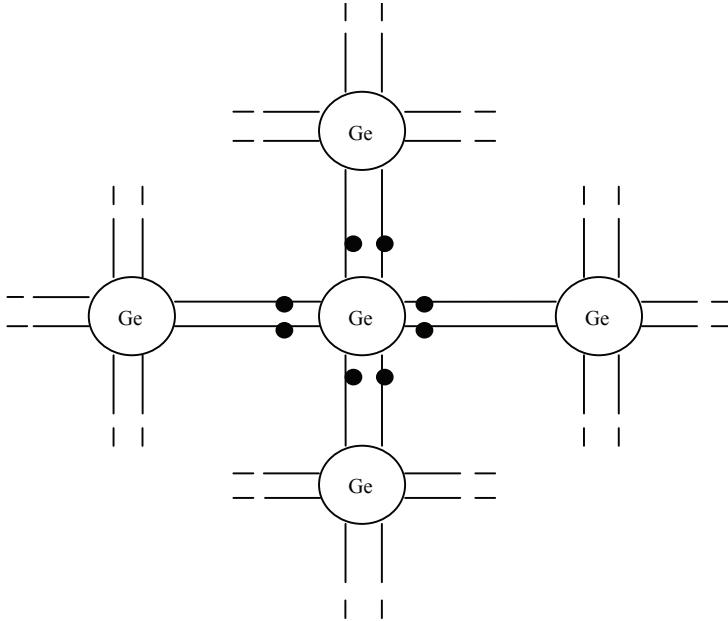


Рис. 3. Плоскостная схема кристаллической решетки германия

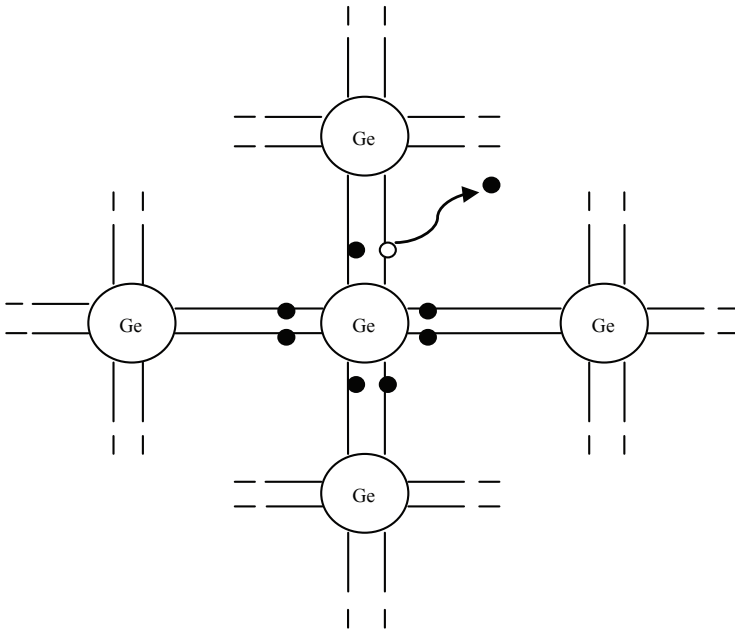


Рис. 4. Возникновение пары электрон — дырка в плоскостной модели полупроводника

Один из электронов, участвующих в ковалентной связи, получив дополнительную энергию, становится электроном проводимости, т. е. свободным носителем заряда, и может перемещаться в кристаллической решетке, а его прежнее место теперь свободно. Именно оно и является дыркой, изображенной на рисунке светлым кружком.

Генерация — это явление возникновения пар электрон — дырка при получении атомом полупроводника дополнительной энергии (тепловой, электрической, световой и пр.), превышающей энергию связи электрона с атомом.

Одновременно с процессом генерации протекает процесс рекомбинации носителей, при котором электрон восстанавливает ковалентную связь, таким образом, рекомбинация — явление исчезновения электрона проводимости и дырки проводимости при их встрече.

Совокупность электронов в твердом теле можно уподобить электронному газу, который заключен в «сосуд», образованный внешними гранями кристалла. Наличие в этом «сосуде» множества неподвижных атомов — узлов решетки — приводит к тому, что свойства электронов в твердом теле отличаются от их свойств в свободном пространстве — вакууме. Одно из важнейших отличий состоит в том, что масса электрона в кристалле не совпадает с его массой m в вакууме. Поэтому в теории твердого тела пользуются понятием эффективной массы m^* , которая зависит от ряда факторов (полная энергия, направление движения и др.). В табл. 1 приведены относительные эффективные массы носителей, приведенные к массе свободного электрона.

Таблица 1

Эффективные массы носителей при комнатной температуре

Эффективная масса	Германий	Кремний	Арсенид галлия	Антимонид индия
m_n^*/m_0 электронов	0,22	0,33	0,07	0,013
m_p^*/m_0 дырок	0,39	0,55	0,50	0,60

Полупроводники обладают электронной и дырочной электропроводностью, которые не наблюдаются в металлах.

При дырочной электропроводности под влиянием приложенной разности потенциалов перемещаются дырки, что эквивалентно пере-

мещению положительных зарядов. Такой процесс показан на рис. 5, на котором изображено для различных моментов времени несколько атомов, расположенных вдоль полупроводника.

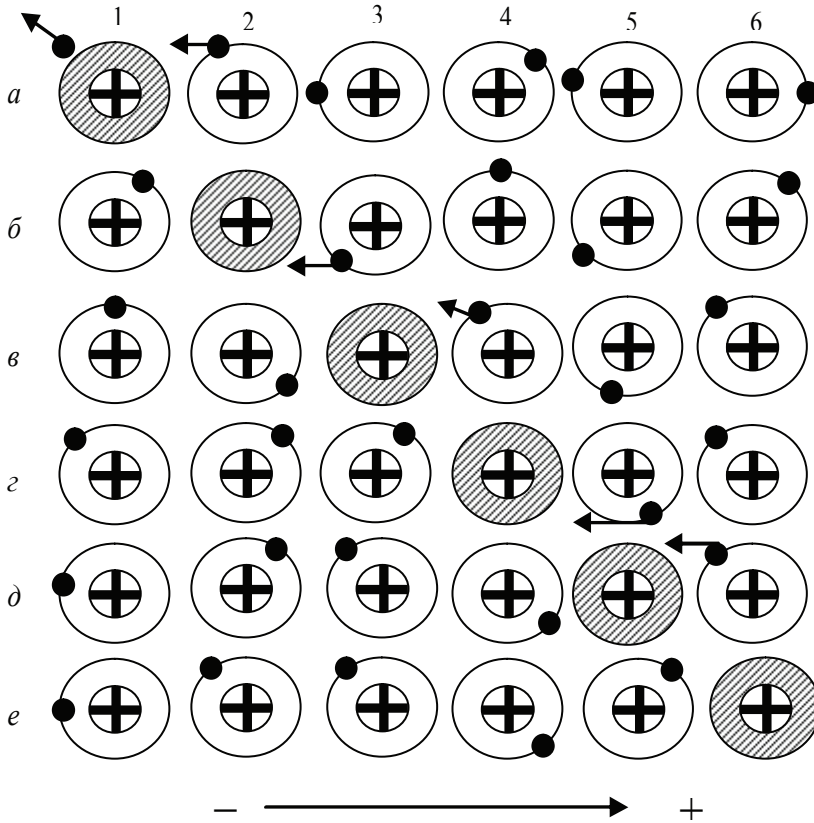


Рис. 5. Принцип дырочной электропроводности

Пусть в начальный момент времени (рис. 5, а) в крайнем атоме слева (1) появилась дырка вследствие того, что из атома ушел электрон. Атом с дыркой (заштрихован) имеет положительный заряд и может притянуть к себе электрон из соседнего атома 2. Если в полупроводнике действует электрическое поле (разность потенциалов), то это поле стремится двигать электроны в направлении от отрицательного потенциала к положительному. Поэтому в следующий момент (рис. 5, б) из атома 2 один электрон перейдет в атом 1 и заполнит дырку, а новая дырка образуется в атоме 2. Далее один электрон из атома 3 пере-

идет в атом 2 и заполнит в нем дырку. Тогда дырка возникнет в атоме 3 (рис. 5, в). Такой процесс будет продолжаться, и дырка перейдет из крайнего левого атома в крайний правый. Иначе говоря, первоначально возникший в атоме 1 положительный заряд перейдет в атом 6 (рис. 5, е).

Таким образом, при дырочной электропроводности в действительности тоже перемещаются электроны, но более ограниченно, чем при электронной электропроводности. Электроны переходят из данных атомов только в соседние. Результатом этого является перемещение положительных зарядов — дырок — в направлении, противоположном движению электронов.

Электроны и дырки, которые могут перемещаться и поэтому создавать электропроводность, называют подвижными носителями заряда, или носителями заряда.

Итак, с точки зрения модели ковалентной связи, дырка — это разорванная ковалентная связь, ведущая себя, как подвижный носитель заряда, равный по модулю заряду электрона.

Будем обозначать концентрации (их число в единице объема, например, в 1 см^3) электронов и дырок проводимости соответственно буквами n и p (от слов *negative* — отрицательный и *positive* — положительный). Очевидно, что для собственного полупроводника электрон проводимости и дырка проводимости всегда образуются парами, следовательно, $n_i = p_i$. Индекс i здесь указывает, что эти концентрации относятся к собственному полупроводнику.

В собственном полупроводнике осуществляется электронная и дырочная электропроводность. Следовательно, электрический ток в собственном полупроводнике определяется двумя составляющими: электронным и дырочным токами, текущими в одном направлении.

Модель энергетических зон

Электроны в твердом теле не могут обладать произвольными энергиями. Энергия каждого электрона может принимать лишь определенные значения, называемые уровнями энергии, или энергетическими уровнями.

Электроны, расположенные ближе к ядру атома, обладают меньшими энергиями, т. е. находятся на более низких энергетических уровнях. Чтобы удалить электрон от ядра, надо преодолеть взаимное притяже-

ние между электроном и ядром. Для этого надо затратить некоторую энергию. Поэтому удаленные от ядра электроны обладают большими энергиями, т. е. они находятся на более высоких энергетических уровнях.

Твердое тело представляет собой множество атомов, сильно взаимодействующих благодаря малым межатомным расстояниям. Поэтому всю совокупность атомов в куске твердого тела следует рассматривать как единое целое, как гигантскую молекулу, которая подобно атому характеризуется некоторым единым для всего тела энергетическим спектром. Особенность этого спектра в том, что он состоит не из дискретных разрешенных уровней, а из дискретных разрешенных зон. Каждая зона происходит от соответствующего атомного уровня, который как бы расщепляется при сближении атомов. Таким образом, для кристалла с межатомным расстоянием d_0 получается определенная зонная диаграмма, в которой разрешенные зоны чередуются с запрещенными зонами. Ширина тех и других обычно не превышает несколько электронвольт и не зависит от числа атомов в твердом теле, т. е. от его размеров.

Строго говоря, разрешенные зоны имеют дискретную структуру и состоят из стольких уровней, сколько атомов имеется в данном теле. Однако количество атомов даже в микроскопических объемах настолько велико, что энергетические расстояния между уровнями зоны в реальных случаях не превышают 10^{-17} эВ, т. е. разрешенные зоны практически можно считать сплошными.

Проводимость в твердом теле возможна лишь тогда, когда возможен переход электрона на ближайший энергетический уровень. Значит, в проводимости могут участвовать электроны только тех зон, в которых есть свободные уровни. Такие свободные уровни иногда имеются в верхней разрешенной зоне. Поэтому верхнюю зону твердого тела, не заполненную (или не полностью заполненную) электронами при нулевой абсолютной температуре, называют зоной проводимости.

Зону, ближайшую к зоне проводимости, называют валентной. При нулевой температуре она полностью заполнена. Следовательно, электроны такой зоны не могут участвовать в проводимости. Но при температуре отличной от нуля, в верхней части валентной зоны образуются свободные уровни, и эта зона также может обусловить проводимость. Таким образом, все существенные процессы в полупроводниковых приборах можно изучить, рассматривая только две смежные

зоны: зону проводимости и валентную. Разрешенные энергетические зоны разделены интервалами энергий, которыми электроны не могут обладать и которые называются запрещенными зонами.

Зонная структура твердого тела при нулевой температуре лежит в основе классификации металлов, полупроводников и диэлектриков. У металлов зона проводимости и валентная зона взаимно перекрываются (рис. 6, а), поэтому даже при нулевой температуре в зоне проводимости находится значительное количество электронов и, следовательно, имеет место проводимость. У полупроводников и диэлектриков при нулевой температуре зона проводимости пуста и проводимость отсутствует, в этом их качественное отличие от металлов (рис. 6, б). Различия же между полупроводниками и диэлектриками в основном количественные и обусловлены значительно большей шириной запрещенной зоны.

Вещества, у которых ширина запрещенной зоны $\Delta W_3 \leq 3$ эВ, относятся к полупроводникам, ΔW_3 , эВ: германий — 0,72; кремний — 1,12; арсенид галлия — 1,41; антимонид индия — 0,18, а при $\Delta W_3 > 3$ эВ — к диэлектрикам. Но и этот критерий классификации веществ неоднозначен (алмаз с $\Delta W_3 \approx 7$ эВ используется в полупроводниковой технике).

Поскольку зона проводимости практически сплошная, энергия электронов в этой зоне может меняться непрерывно, как у изолированных электронов в вакууме, постольку электроны в зоне проводимости называются свободными. Термин «свободный» характеризует возможность перемещения электрона внутри твердого тела.

При нулевой абсолютной температуре, проводимость в собственном полупроводнике отсутствует, потому что зона проводимости пуста, а валентная зона заполнена. При любой температуре отличной от нуля, в кристалле появляются фононы, энергетический спектр которых непрерывен. Фононы с энергией, превышающей ширину запрещенной зоны, переводят некоторые электроны из верхней части валентной зоны в зону проводимости. В результате этого в зоне проводимости появляются свободные электроны, а в валентной зоне — незаполненные уровни (рис. 7). Те и другие образуются одновременно и в равных количествах. Теперь электроны обеих зон могут двигаться в электрическом поле, обеспечивая проводимость кристалла.

Исходя из модели энергетических зон дырка — это незаполненный уровень валентной зоны.

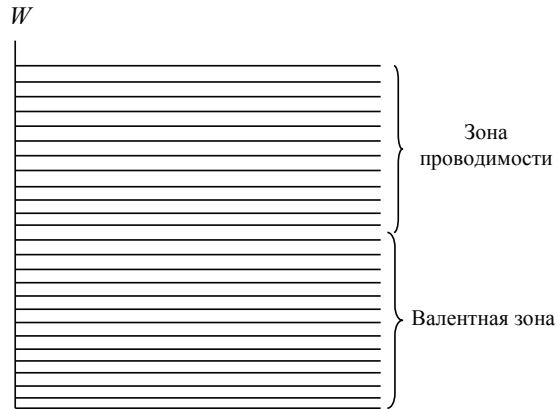
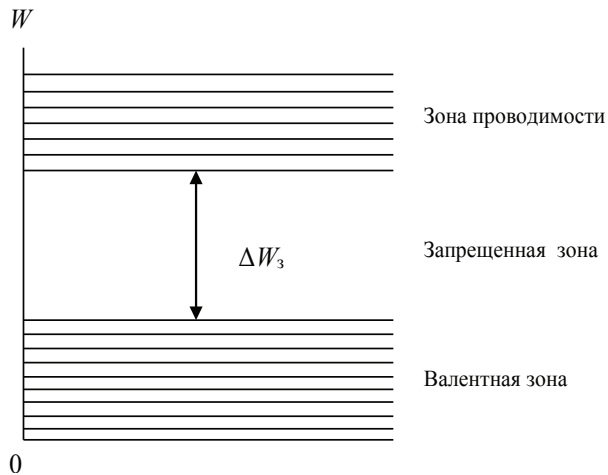
a*б*

Рис. 6. Схема зонной диаграммы для металла (*a*) и диэлектрика (*б*)

Поясним понятие дырки с точки зрения зонной теории. Для этого предположим, что к полупроводнику приложено внешнее напряжение. При наличии электрического поля приходят в движение не только свободные электроны зоны проводимости, но и вся совокупность электронов валентной зоны (поскольку в ней имеются незаполненные верхние уровни, т.е. отсутствует часть электронов с отрицательной

эффективной массой). Оказывается, что движение огромного количества электронов валентной зоны с самыми различными эффективными массами эквивалентно движению ограниченного (равного числу незаполненных уровней) количества фиктивных квазичастиц с положительным зарядом и положительной эффективной массой. Именно эти квазичастицы получили название дырок.

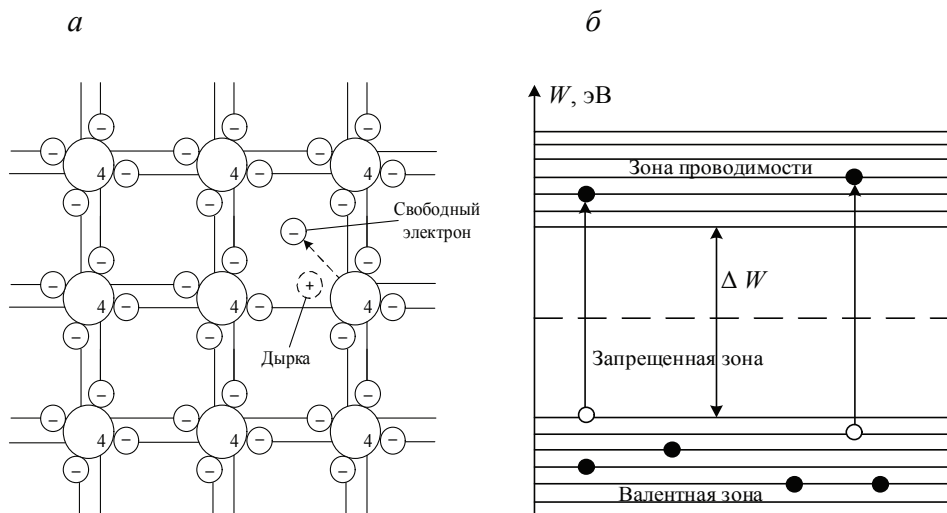


Рис. 7. Условное изображение кристаллической решетки (а) и энергетическая диаграмма (б) полупроводника с собственной электропроводностью

Целесообразность понятие дырок состоит в том, что весьма сложный анализ кинетики электронов в почти полной валентной зоне сводится к сравнительно простому анализу кинетики дырок в почти пустой (по отношению к дыркам) валентной зоне. При этом поведение дырок оказывается аналогичным поведению свободных электронов в почти пустой (по отношению к электронам) зоне проводимости. Разница сводится к различию эффективных масс обоих типов носителей зарядов.

Следует подчеркнуть, что понятие эффективной массы удобно и корректно лишь по отношению к кинетике носителей; его нельзя отождествлять с понятием массы в обычном смысле этого слова. Так, различие эффективных масс отнюдь не характеризует различие силы тяжести.

Итак, перевод электрона из валентной зоны в зону проводимости аналогичен возбуждению электрона в изолированном атоме, т. е. переводу его на один из верхних уровней (генерация). Такое возбужденное состояние, как известно, является временным: через некоторое время электрон в атоме возвращается на исходный уровень, а в твердом теле — в валентную зону. В корпускулярном аспекте это возвращение есть рекомбинация электрона и дырки. Среднее время возбужденного состояния атома (т. е. пребывания его в зоне проводимости) называется временем жизни.

Очевидно, что количество электронно-дырочных пар тем больше, чем выше температура и меньше ширина запрещенной зоны. Из-за процессов генерации и рекомбинации носителей зарядов, при данной температуре устанавливается определенная концентрация электронов в зоне проводимости n_i и равная ей концентрация дырок p_i в валентной зоне.

Скорость генерации $V_{\text{ген}}$ — число электронно-дырочных пар, возникающих в единицу времени, — зависит от температуры и ширины запрещенной зоны.

Скорость рекомбинации $V_{\text{рек}}$ — число электронно-дырочных пар, исчезающих в единицу времени, — зависит от концентрации носителей и свойств полупроводника

$$V_{\text{рек}} = r n_i p_i,$$

где r — коэффициент пропорциональности.

Для собственного полупроводника выполняется $n_i = p_i$.

Для состояния термодинамического равновесия ($T = \text{const}$) обязательно выполняется равенство этих скоростей

$$V_{\text{ген}} = V_{\text{рек}} = r n_i p_i = r n_i^2.$$

В процессе генерации электронно-дырочных пар и их рекомбинации в собственном полупроводнике, для каждого значения температуры устанавливается равновесная (или собственная) концентрация носителей n_i .

Равновесная концентрация носителей заряда

Как определить концентрацию носителей в заданном полупроводнике при заданной температуре? Это — важнейшая задача статистической физики. Для решения данной задачи необходимо знать число квантовых состояний в заданном интервале энергий и вероятность

нахождения частиц в этих состояниях. Следовательно, для определения концентрации носителей заряда в полупроводнике необходимо знать фактическое число состояний, занятых электронами и дырками (графическая интерпретация расчета концентрации носителей заряда в собственном полупроводнике представлена на рис. 8, а на рис. 8, I — энергетическая (зонная) диаграмма собственного полупроводника).

В теории твердого тела показывается, что энергетические уровни распределены по высоте разрешенной зоны неравномерно: плотность их меняется от границы в глубь зоны.

Обозначим через $N(W)$ — плотность состояний, т. е. число квантовых состояний в единичном интервале энергии для единичного объема кристалла (рис. 8, II). Наиболее простым частным случаем будет идеальный (безпримесный), другими словами, собственный кристалл полупроводника, находящийся при температуре абсолютного нуля $T = 0$. В этом случае все возможные уровни в валентной зоне будут заполнены электронами, в зоне проводимости нет электронов и распределение электронов по энергиям будет строго соответствовать закону изменения плотности разрешенных уровней (на рис. 8, II это распределение заштриховано).

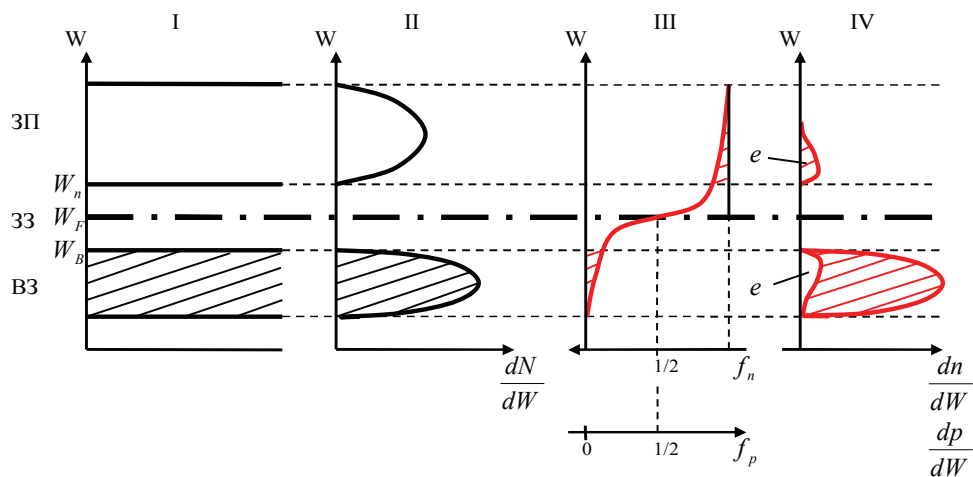


Рис. 8. Графическая интерпретация расчета концентрации носителей заряда в собственном полупроводнике

С ростом температуры некоторое количество электронов покидает валентную зону и переходит на уровни зоны проводимости. Возникает

вопрос: какова вероятность того, что тот или иной уровень валентной зоны лишится электрона и превратится в дырку? Ответ на этот вопрос дает статистика Ферми–Дирака.

Согласно статистике Ферми–Дирака вероятность того, что состояние с энергией W при заданной температуре T занято электроном, выражается функцией Ферми для электронов

$$f_n(W, T) = \frac{1}{\exp\left(\frac{W - W_F}{kT}\right) + 1},$$

где W_F — энергия, называемая уровнем Ферми; k — постоянная Больцмана; T — абсолютная температура.

Вероятность выражается в долях единицы (см. рис. 8, III). Имеется только две возможности: уровень занят электроном или уровень не занят электроном. Вероятность того, что уровень валентной зоны не занят электроном, есть вероятность нахождения на данном уровне дырки. Сумма вероятностей этих событий должна быть равна 1: $f_n + f_p = 1$ (масштаб осей f_n и f_p на рис. 8 III). Функция Ферми для дырок

$$f_p(W, T) = \frac{1}{\exp\left(\frac{W_F - W}{kT}\right) + 1}.$$

При $T = 0$ функция Ферми имеет вид ступеньки и показывает, что все энергетические состояния, превышающие уровень Ферми, свободны, а все энергетические состояния ниже уровня Ферми заняты.

$$T = 0 \begin{cases} f_n(W_0) = 0, & W > W_F \\ f_n(W_0) = 1, & W < W_F \end{cases}.$$

С ростом температуры, функция Ферми превращается в плавную (симметричную) кривую (см. рис. 8, III). Вероятность нахождения электрона в зоне проводимости становится отличной от нуля, а в валентной зоне становится отличной от нуля вероятность появления свободных уровней — дырок. Поскольку эта кривая симметрична, эти вероятности равны.

Количество электронов, занимающих разрешенные уровни в некоторой элементарной полосе ∂W с плотностью уровней $N(W)$, определяется числом разрешенных уровней в данной элементарной полосе $N(W) \cdot \partial W$ и вероятностью их заполнения f_n

$$\partial n = N(W) \cdot f_n(W) \partial W.$$

Следовательно, количество электронов, находящихся в зоне проводимости, будет

$$n = \int_{W_{II}}^{\infty} N(W) \cdot f_n(W) dW,$$

а число дырок в валентной зоне составит

$$p = \int_0^{W_B} N(W) \cdot f_p(W) dW.$$

После преобразований можно получить следующий результат: концентрация электронов n_i будет определяться по выражению

$$n_i = \sqrt{N_c N_v} \exp\left(-\frac{W_{II} - W_F}{kT}\right) \approx N_c \exp\left(-\frac{\Delta W_3}{2kT}\right),$$

а концентрация дырок p_i —

$$p_i = \sqrt{N_c N_v} \exp\left(-\frac{W_F - W_B}{kT}\right) \approx N_v \exp\left(-\frac{\Delta W_3}{2kT}\right),$$

где N_c — эффективная плотность разрешенных состояний в зоне проводимости, $N_c = 2 \left(\frac{2\pi m_n^* kT}{h^2} \right)^{3/2}$, а N_v — эффективная плотность разре-

шенных состояний валентной зоны, $N_v = 2 \left(\frac{2\pi m_p^* kT}{h^2} \right)^{3/2}$. Здесь m_n^*

и m_p^* — эффективные массы электрона и дырки соответственно; h — постоянная Планка.

Если подставить численные значения универсальных констант, то будем иметь $N_c = 2 \left(\frac{2\pi m_n^* kT}{h^2} \right)^{3/2} = 4,82 \cdot 10^{15} (m_n^*/m_0)^{3/2} T^{3/2}$.

Если положить $m_n^* = m_p^* = m_0$, то $N_c = N_v \approx 2,5 \cdot 10^{19} \text{ см}^{-3}$ при $T = 300 \text{ К}$. Это составляет приблизительно 1/2000 числа атомов в 1 см^3 .

Распределение электронов и дырок по энергиям представлено на рис. 8, IV, из которого видно, что эти распределения носителей за-

ряда по энергиям носят экспоненциальный характер. С ростом энергии, число электронов в зоне проводимости быстро уменьшается. При $T = 300$ К основная часть электронов проводимости имеет энергию близкую к энергии дна зоны проводимости, дырки проводимости имеют энергию близкую к энергии потолка валентной зоны.

Равновесная концентрация носителей заряда зависит от температуры

$$n_i \sim T^{3/2} \exp\left(-\frac{\Delta W_3}{2kT}\right),$$

причем экспонента является определяющим множителем в этом выражении. Приблизительно можно считать, что n_i будет возрастать на 5–7 % при увеличении температуры на 1 градус.

Равновесная концентрация носителей заряда зависит от ширины запрещенной зоны полупроводника: чем больше ширина запрещенной зоны, тем меньше равновесная концентрация n_i , см⁻³ (ниже приведена при комнатной температуре) — для германия — $2,5 \cdot 10^{13}$; кремния — $1,4 \cdot 10^{10}$, арсенида галлия — $1,5 \cdot 10^6$.

Из всего выше сказанного можно сделать следующие выводы:

- 1) собственный полупроводник — беспримесный полупроводник;
- 2) дырка — это разорванная ковалентная связь, ведущая себя как подвижный носитель заряда, равный по модулю заряду электрона, или незаполненный уровень валентной зоны, согласно модели энергетических зон;
- 3) энергетический спектр твердого тела состоит из дискретных разрешенных зон. Все существенные процессы в полупроводниковых приборах можно изучить, рассматривая только две смежные зоны — зону проводимости и валентную;
- 4) при $T = 300$ К основная часть электронов имеет энергию близкую к энергии дна зоны проводимости, а дырок — энергию близкую к энергии потолка валентной зоны;
- 5) распределения носителей заряда по энергиям носят экспоненциальный характер;
- 6) равновесная концентрация носителей заряда зависит от ширины запрещенной зоны полупроводника — чем больше ширина запрещенной зоны, тем меньше равновесная концентрация;
- 7) равновесная концентрация носителей заряда экспоненциально зависит от температуры.

2.2. Электронный полупроводник

Определение примесного полупроводника

Примесный — это такой полупроводник, в котором для изменения его электрофизических свойств нарушена структура кристаллической решетки.

Большинство полупроводниковых приборов изготавливают на основе примесных полупроводников. Чтобы превратить собственный полупроводник в примесный, необходимо ввести в его кристаллическую решетку некоторое количество специально подобранной химической добавки, т. е. осуществить легирование полупроводника. Если примесные атомы находятся в узлах кристаллической решетки, то их называют примесями замещения, если в междоузлиях — примесями внедрения. На практике легирование осуществляется при помощи процесса диффузии.

Роль примесей могут играть и всевозможные дефекты структуры. К числу таких дефектов относятся вакансии и атомы, расположенные между узлами кристаллической решетки.

В электронике применяются полупроводниковые материалы, у которых концентрация примеси составляет $10^{14} - 10^{18} \text{ см}^{-3}$.

Получение полупроводников с таким малым и строго дозированным содержанием нужной примеси является весьма сложным процессом. При этом исходный полупроводник, к которому добавляется примесь, должен быть очень чистым. Для германия посторонние примеси допускаются в количестве примерно не более $10^{-8}\%$, т. е. не более одного атома на 10 миллиардов атомов германия. А для кремния посторонних примесей допускается еще меньше: они не должны превышать $10^{-11}\%$.

Модель ковалентной связи

Рассмотрим, как влияет валентность примесных атомов замещения на характер их поведения в ковалентных полупроводниках типа кремния и германия. Предположим, что в кристаллической решетке кремния часть основных атомов замещена атомами мышьяка As (элемента V подгруппы периодической таблицы Менделеева), у которого на внешней оболочке расположено пять валентных электронов. Встра-

иваясь в узел решетки, атом мышьяка отдает четыре электрона на образование связей с ближайшими соседями, а пятый электрон оказывается лишним, т. е. не участвует в создании ковалентных связей. Из-за большой диэлектрической проницаемости среды (полупроводника), кулоновское притяжение этого лишнего электрона ядром в значительной мере ослаблено. Поэтому радиус электронной орбиты оказывается большим, охватывает несколько межатомных расстояний. Достаточно небольшого теплового возбуждения, чтобы оторвать избыточный электрон от примесного атома. Для этого необходимо затратить энергию равную $\Delta W_{\text{д}}$. Атом примеси, потерявший электрон, превращается в положительно заряженную частицу, неподвижно закрепленную в данном месте решетки (рис. 9).

Таким образом, примеси замещения, валентность которых превышает валентность основных атомов решетки, проявляют свойство доноров. Кроме мышьяка типичными донорами в кремнии и германии являются фосфор Р и сурьма Sb.

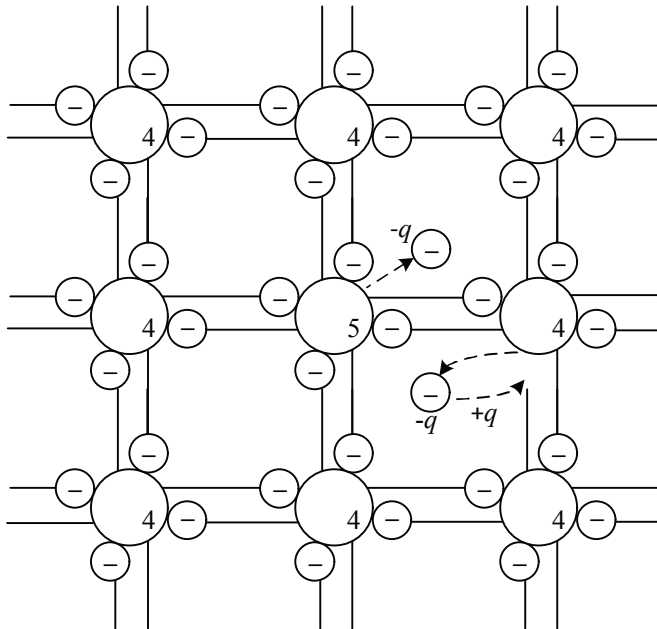


Рис. 9. Замещение примесными атомами основных атомов решетки

Поскольку избыточный электрон примесного атома движется по круговой орбите в кулоновском поле однократного заряженного

положительного иона, донор можно рассматривать как водородоподобный атом, помещенный в среду с диэлектрической проницаемостью ϵ . Поэтому для оценки энергии ионизации доноров можно использовать формулу

$$\Delta W_{\text{д}} = \frac{q^4 m_n^*}{8\epsilon_0^2 \epsilon^2 h^2},$$

где m_n^* — эффективная масса электрона. Отсюда

$$\Delta W_{\text{д}} = W_0 \frac{1}{\epsilon^2} \left(\frac{m_n}{m_0} \right),$$

где W_0 — энергии ионизации атома водорода.

Для германия $\epsilon = 16$, для кремния $\epsilon = 12$, поэтому энергия ионизации примеси в них соответственно в 256 и 144 раз меньше энергии ионизации атома водорода. Поскольку m_n/m_0 несколько меньше единицы, можно заключить, что энергия ионизации донора в германии составляет 0,05 эВ, а в кремнии — менее 0,10 эВ.

Свободные электроны примесного происхождения добавляются к собственным свободным электронам. Поэтому проводимость полупроводника становится преимущественно электронной.

Модель энергетических зон

Примеси создают дополнительные уровни в запрещенной зоне полупроводника. При малой концентрации примесей, расстояние между примесными атомами велико, их электронные оболочки не взаимодействуют друг с другом. Вследствие этого примесные энергетические уровни являются дискретными, т. е. не расщепляются в зону, как это имеет место для уровней основных атомов кристаллической решетки (рис. 10).

Примесные уровни, заполненные электронами, при отсутствии внешних энергетических воздействий, расположены в запрещенной зоне вблизи нижнего края зоны проводимости. При внешнем возбуждении электроны с примесных уровней могут легко переходить в свободную зону и участвовать в процессе электропроводности. Энергия, необходимая для таких переходов, значительно меньше энергии ионизации собственных атомов полупроводника, т. е. ширины запрещенной зоны

$$\Delta W_{\text{д}} = W_{\text{п}} - W_{\text{д}} \ll \Delta W_{\text{з}}.$$

Примеси, поставляющие электроны в зону проводимости полупроводника, называют донорами. А энергетический уровень, расположенный вблизи дна зоны проводимости, содержащий электрон в невозбужденном состоянии и способный отдать электрон в зону проводимости при возбуждении, называется донорным уровнем.

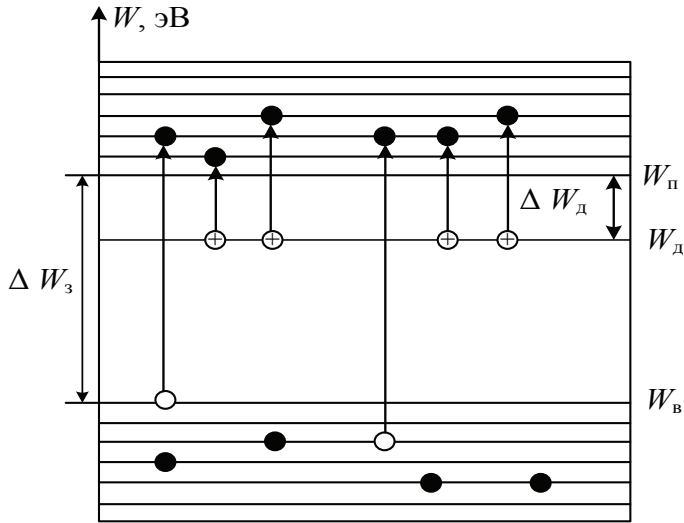


Рис. 10. Энергетическая диаграмма электронного полупроводника

При относительно невысоких температурах, переходы электронов из валентной зоны в зону проводимости не играют существенной роли. В таких материалах концентрация электронов превышает концентрацию дырок, вследствие чего они получили название полупроводников *n*-типа.

Расчет равновесной концентрации носителей заряда

Поскольку в примесных полупроводниках концентрации электронов и дырок резко различны, принято называть носителей преобладающего типа основными, а носителей другого типа неосновными. В полупроводнике *n*-типа основными носителями являются электроны (n_n), а неосновными носителями — дырки (p_n), причем $n_n = N_{\text{д}} + n_i \approx N_{\text{д}}$ и $p_n \ll n_i$.

Для доказательства этого соотношения необходимо знать число квантовых состояний в заданном интервале энергий и вероятность нахождения частиц в этих состояниях. Следовательно, для определения концентрации носителей заряда в полупроводнике необходимо знать фактическое число состояний, занятых электронами и дырками. Расчет равновесной концентрации носителей приведен для собственного полупроводника. Задача расчета равновесной концентрации носителей для электронного полупроводника аналогична, но необходимо учесть изменение уровня Ферми.

В полупроводнике n -типа большое количество электронов переходит с донорных уровней. Поскольку появление таких электронов в зоне проводимости не связано с появлением дырок в валентной зоне, вероятность появления электронов в зоне проводимости должна быть больше вероятности появления дырок в валентной зоне. Очевидно, это будет возможно в том случае, если уровень Ферми сместится в направлении дна зоны проводимости W_{Π} .

Графическая интерпретация расчета концентрации носителей заряда в полупроводнике n -типа представлена на рис. 11.

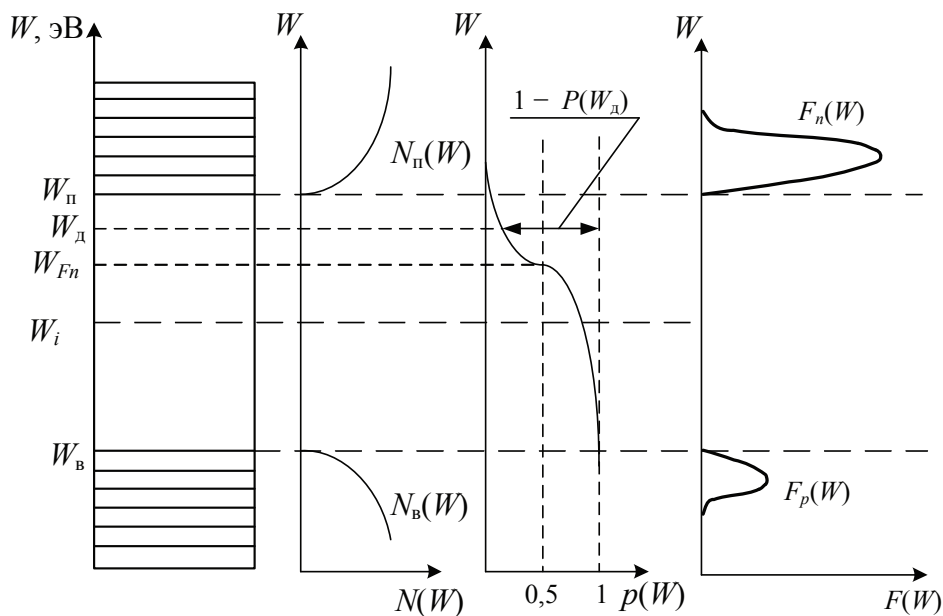


Рис. 11. Графическая интерпретация расчета концентрации носителей заряда в полупроводнике n -типа

Площадь под графиком $F_n(W)$ пропорциональна концентрации электронов n_n , а площадь под графиком $F_p(W)$ — концентрации дырок p_n .

Если считать, что при комнатной температуре все атомы донорных примесей ионизированы ($n_n = N_d$, $p_n \approx 0$), уровень Ферми n -типа можно записать

$$W_{Fn} = W_{\Pi} - kT \ln \frac{N_c}{N_d},$$

где N_c — эффективная плотность разрешенных уровней в зоне проводимости.

Из данного уравнения следует, что увеличение концентрации примеси приближает уровень Ферми к границам запрещенной зоны. При концентрации примесей порядка $10^{15} - 10^{19} \text{ см}^{-3}$, уровень Ферми расположен сравнительно далеко от границ запрещенной зоны. Такое состояние полупроводника называется невырожденным. При более высокой концентрации примесей возрастает взаимодействие примесных атомов и происходит расширение полосы, занимаемой энергетическими уровнями этих атомов, в результате эта полоса сливается с ближайшей к ней зоной разрешенных уровней, а уровень Ферми оказывается за пределами запрещенной зоны. Такое состояние полупроводника называется вырожденным.

Уравнения для расчета концентрации носителей заряда в электронном полупроводнике с учетом сдвига уровня Ферми относительно середины запрещенной зоны легко приводятся к виду

$$n_n = n_i \exp\left(\frac{W_{Fn} - W_{Fi}}{kT}\right);$$

$$p_n = n_i \exp\left(\frac{W_{Fi} - W_{Fn}}{kT}\right),$$

где n_i и W_{Fi} — равновесная концентрация и уровень Ферми в собственном полупроводнике.

Из данного выражения следует важное соотношение $n_n p_n = n_i^2$, суть которого состоит в том, что увеличение концентрации основных носителей заряда за счет увеличения концентрации примесей сопровождается уменьшением концентрации неосновных носителей заряда. Это — закон действующих масс: в невырожденном полупроводнике произведение концентраций свободных электронов и дырок при тер-

модинамическом равновесии есть постоянная величина, равная квадрату равновесной концентрации при данной температуре. Следовательно, концентрация носителей зависит от положения уровня Ферми. В электронном полупроводнике концентрация электронов в основном обусловлена переходом с энергетических уровней W_d на энергетические уровни зоны проводимости. Поэтому при комнатной температуре концентрация основных носителей n_n должна быть равна концентрации ионизированных доноров $n_n \approx N_d$. Концентрацию неосновных носителей заряда можно найти следующим образом: $p_n = \frac{n_i^2}{n_n}$.

Закон действующих масс позволяет рассчитать концентрации основных и неосновных носителей заряда в заданном полупроводнике при известной температуре и степени легирования. Из закона действующих масс следует, что при комнатной температуре концентрация неосновных носителей очень мала. Так, если $N_d = 1 \cdot 10^{17} \text{ см}^{-3}$, то при комнатной температуре в кремнии концентрация дырок определяется по формуле

$$p_n = \frac{n_i^2}{n_n} = \frac{n_i^2}{N_d} = \frac{(1,4 \cdot 10^{10})^2}{(1 \cdot 10^{17})} \approx 2 \cdot 10^3 \text{ см}^{-3}$$

и составляет всего 2000 см^{-3} , т. е. на 14 порядков меньше, чем концентрация электронов. А для германия, при той же степени легирования и температуре, концентрация дырок составит

$$p_n = \frac{n_i^2}{N_d} = \frac{(2,5 \cdot 10^{13})^2}{(1 \cdot 10^{17})} \approx 6,25 \cdot 10^9 \text{ см}^{-3},$$

т. е. дырок (неосновных носителей) на 8 порядков меньше, чем электронов, но в германии неосновных носителей заряда будет на 6 порядков больше, чем в кремнии.

Рассмотрим зависимость концентрации носителей от температуры (рис. 12). Разобьем график на три участка. Первый — участок ионизации примесей. Наблюдается рост концентрации основных носителей. Второй — участок примесной проводимости: концентрация основных носителей практически постоянна ($n_n \approx \text{const}$). Однако, с повышением температуры, концентрация неосновных носителей возрастает очень резко — пропорционально n_i^2 , т. е. несравненно быстрее, чем даже концентрация собственных носителей. Так, в кремнии рост температуры на 50°C сопровождается увеличением концентрации неосновных но-

сителей примерно на 3 порядка. $n_n \gg p_n$ и $n_n \gg n_i$. Третий — участок собственной проводимости $n_n \approx p_n$. На рис. 12 выделен диапазон рабочих температур $T_{\text{раб}}$. Обычно он составляет $-60^\circ\text{C} \dots +60^\circ\text{C}$. В этом диапазоне концентрация основных носителей $n_n = N_d$.

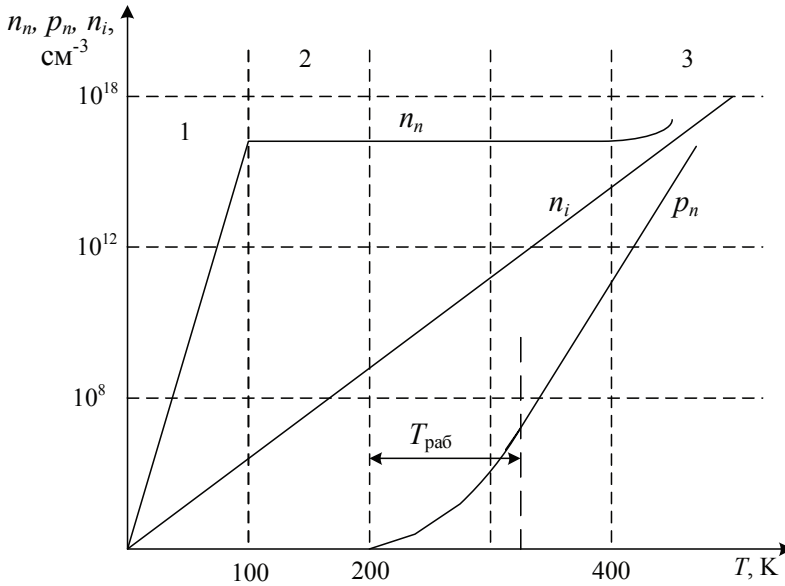


Рис. 12. Зависимость концентрации носителей от температуры для полупроводника n -типа

Температурный диапазон второго участка ограничен снизу температурой полной ионизации примесей (для кремния $-100 \dots -70^\circ\text{C}$), а сверху — критической температурой, при которой примесный полупроводник теряет примесный характер проводимости.

Поясним закон действующих масс с точки зрения зонной теории. Концентрация неосновных носителей в примесном полупроводнике уменьшается во столько раз, во сколько увеличивается концентрация основных носителей. Выше, в примере, было получено: если в германии i -типа $n_i = p_i = 2,5 \cdot 10^{13} \text{ см}^{-3}$, а после добавления донорной примеси, концентрация электронов возросла в 10^4 раз и стала $n_n = 10^{17} \text{ см}^{-3}$, то концентрация неосновных носителей (дырок) уменьшится в 10^4 раз и станет $p_n = 10^9 \text{ см}^{-3}$, т. е. будет в 10^8 раз меньше концентрации основных носителей. Это объясняется тем, что при увеличении в 10^4 раз концентрации электронов проводимости, полученных от донорных ато-

мов, нижние энергетические уровни зоны проводимости оказываются занятыми и переход электронов из валентной зоны возможен только на более высокие уровни зоны проводимости. Но для такого перехода электроны должны иметь большую энергию, чем в собственном полупроводнике, и поэтому значительно меньшее число электронов может его осуществить. Соответственно значительно уменьшается число дырок проводимости в валентной зоне. Оказывается, что всегда для примесного полупроводника n -типа справедливо соотношение

$$n_n p_n = n_i p_i = n_i^2 = p_i^2.$$

В нашем примере получилось: $10^{17} \cdot 10^9 = (10^{13})^2 = 10^{26}$.

Рассмотренные примеры наглядно показывают, что ничтожно малое количество примеси существенно изменяет характер электропроводности и проводимость полупроводника. Действительно, концентрация примеси 10^{17} см^{-3} при числе атомов германия $4,4 \cdot 10^{22}$ в 1 см^3 означает, что добавляется всего лишь 1 атом примеси на 440 тыс. атомов германия, т. е. примесь составляет менее $10^{-3}\%$. Но в результате этого концентрация основных носителей возрастает в 10 000 раз и соответственно увеличивается электронная проводимость.

Из вышесказанного можно сделать следующие выводы:

- 1) концентрация как основных, так и неосновных носителей заряда зависит от положения уровня Ферми;
- 2) введение в полупроводник примесей сдвигает уровень Ферми относительно середины запрещенной зоны в электронном полупроводнике вверх к зоне проводимости;
- 3) повышение температуры полупроводника сдвигает уровень Ферми к середине запрещенной зоны;
- 4) увеличение концентрации примесей увеличивает концентрацию основных носителей заряда — электронов и уменьшает концентрацию неосновных носителей заряда — дырок;
- 5) при равной степени легирования, неосновных носителей заряда существенно меньше в полупроводниках с большей запрещенной зоной, а следовательно, с меньшей собственной концентрацией носителей (например, в кремнии на 6 порядков меньше неосновных носителей, чем в германии);
- 6) Для увеличения диапазона рабочих температур $T_{\text{раб}}$ необходимо:
 - использовать полупроводниковый материал с большей запрещенной зоной;
 - увеличивать степень легирования полупроводника.

2.3. Дырочный полупроводник

Примесный — это такой полупроводник, в котором в целях изменения его электрофизических свойств нарушена структура кристаллической решетки. Дырочный полупроводник является одним из видов примесных полупроводников.

Модель ковалентной связи

Дырочным полупроводником, или полупроводником *p*-типа (от лат. *positive* — положительный), называется полупроводник, в кристаллической решетке которого содержатся примесные трехвалентные атомы, называемые акцепторами. В такой кристаллической решетке одна из ковалентных связей остается незаполненной. Свободную связь примесного атома может заполнить электрон, покинувший одну из соседних связей. При этом примесный атом превращается в отрицательный ион, а на том месте, откуда ушел электрон, возникает дырка.

Если в кристалле четырехвалентного элемента, например кремния, часть атомов замещена атомами трехвалентного элемента, например атомами бора В, то для образования четырех ковалентных связей у примесного атома не хватает одного электрона (рис. 13). Одна из ковалентных связей остается незаполненной, однако эта незаполненная связь не имеет заряда, и атом В и сменный с ним атом Si является электрически нейтральными. В незаполненную связь может перейти электрон от атома основного элемента полупроводника за счет разрыва ковалентной связи. Атомы примеси могут быть ионизированы под действием энергии, поступающей в кристалл в виде квантов света, теплоты и т. д. Во внешней оболочке В при этом появляется лишний электрон, атом В превращается в отрицательный ион.

Нарушается и электрическая нейтральность той связи, откуда электрон перешел в связь атома В. В этой связи появляется положительный подвижный заряд — дырка. Примеси, захватывающие электроны из валентной зоны, называют акцепторными, или акцепторами. Таким образом, примеси замещения, валентность которых меньше валентности основных атомов решетки, проявляют свойство акцепторов. Кроме бора В типичными акцепторами в кремнии и германии являются примеси галлия Ga и индия In.

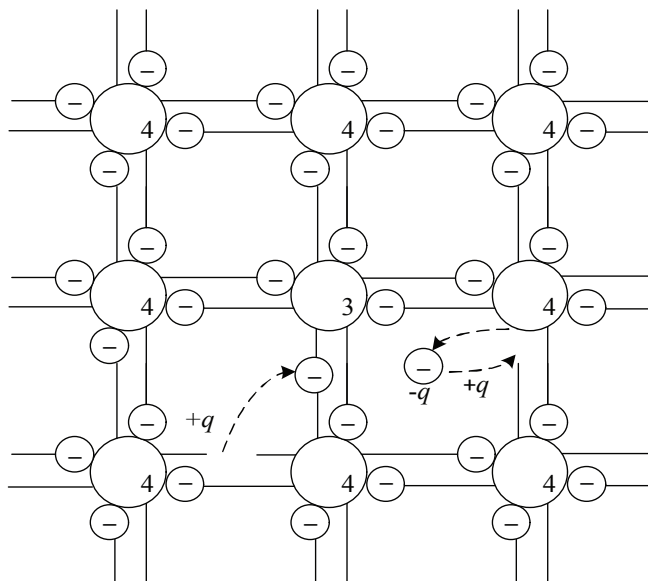


Рис. 13. Условное изображение кристаллической решетки полупроводника с дырочной электропроводностью

Энергия активизации акцепторов W_A составляет для германия 0,01–0,012 эВ и для кремния 0,04–0,16 эВ, что значительно меньше ширины запрещенной зоны беспримесного полупроводника.

Модель энергетических зон

Примеси создают дополнительные уровни в запрещенной зоне полупроводника. При малой концентрации примесей, расстояние между примесными атомами велико, их электронные оболочки не взаимодействуют друг с другом. Вследствие этого примесные энергетические уровни являются дискретными, т. е. не расщепляются в зону, как это имеет место для уровней основных атомов кристаллической решетки. Следовательно, энергетические уровни акцепторных атомов располагаются вблизи валентной зоны (рис. 14).

Энергетический уровень, расположенный вблизи потолка валентной зоны, свободный в невозбужденном состоянии и способный принимать валентный электрон при возбуждении, называется акцепторным уровнем.

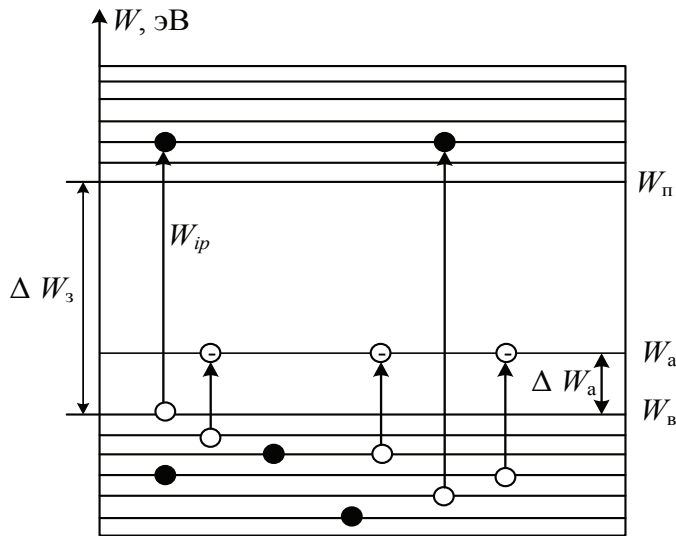


Рис. 14. Энергетическая диаграмма полупроводника с дырочной электропроводностью

Энергия ионизации акцептора — это минимальная энергия, которую необходимо сообщить электрону валентной зоны, чтобы перевести его на акцепторный уровень, определяется по выражению $\Delta W_a = W_a - W_B \ll \Delta W_3$.

Из-за малого значения энергии активизации акцепторов, уже при комнатной температуре электроны из валентной зоны переходят на уровни акцепторов. Такие электроны, превращая примесные атомы в отрицательные ионы, теряют способность перемещаться по кристаллической решетке, а образовавшиеся при этом дырки могут участвовать в создании электрического тока.

Расчет равновесной концентрации носителей заряда

За счет ионизации атомов исходного материала, из валентной зоны часть электронов попадает в зону проводимости. Однако электронов в зоне проводимости значительно меньше, чем дырок в валентной зоне, образующихся в результате ионизации акцепторов. Поэтому дырки в таких полупроводниках являются основными носителями заряда, их концентрация обозначается p_p , а электроны — неосновными носителями заряда с концентрацией n_p . Такие полупроводники носят на-

звание полупроводников с дырочной электропроводностью, или полупроводников p -типа.

Для определения концентрации носителей заряда в полупроводнике p -типа необходимо знать фактическое число состояний, занятых электронами и дырками. Расчет равновесной концентрации носителей был приведен для собственного полупроводника. Задача расчета равновесной концентрации носителей для полупроводника p -типа аналогична этому расчету, но вновь необходимо учесть изменение уровня Ферми.

В полупроводнике p -типа большое количество дырок появляется в валентной зоне за счет акцепторного уровня и не связано с появлением электронов в зоне проводимости ($N_A \gg n_i$), вероятность появления дырок в валентной зоне должна быть больше вероятности появления электронов в зоне проводимости. Очевидно, это будет возможно в том случае, если уровень Ферми сместится в направлении потолка валентной зоны W_B .

Графическая интерпретация расчета концентрации носителей заряда в полупроводнике p -типа представлена на рис. 15.

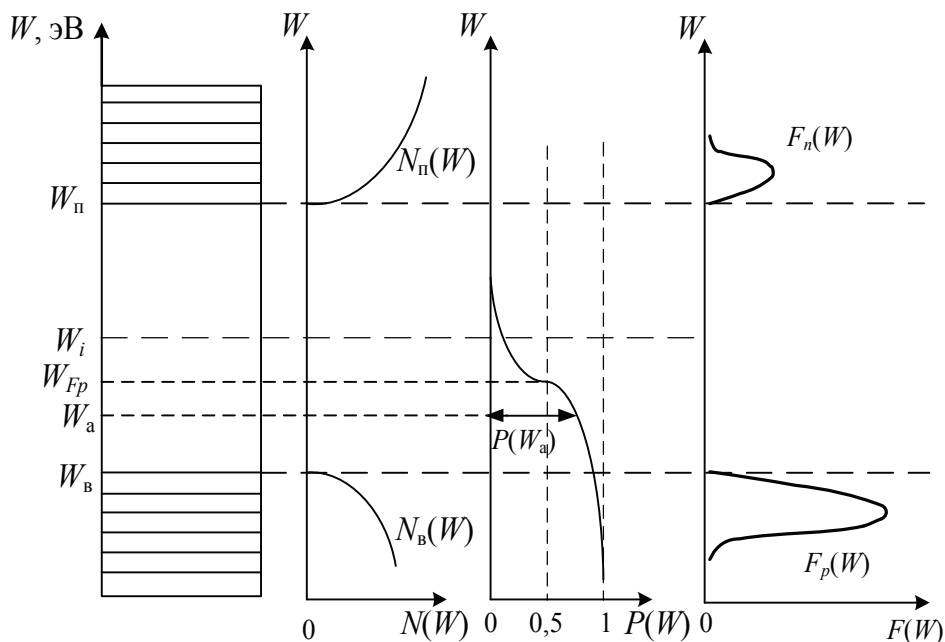


Рис. 15. Графическая интерпретация расчета концентрации носителей заряда в полупроводнике p -типа

Если считать, что при комнатной температуре все атомы акцепторных примесей ионизированы ($p_p = N_A$, $n_p \approx 0$), уровень Ферми n -типа можно записать

$$W_{Fp} = W_B + kT \ln \frac{N_V}{N_A},$$

где N_V — эффективная плотность разрешенных уровней в валентной зоне; N_A — концентрация акцепторных атомов в полупроводнике.

Уравнения для расчета концентрации носителей заряда в дырочном полупроводнике с учетом сдвига уровня Ферми относительно середины запрещенной зоны легко приводятся к виду

$$p_p = n_i \exp\left(\frac{W_{Fi} - W_{Fp}}{kT}\right);$$

$$n_p = n_i \exp\left(\frac{W_{Fp} - W_{Fi}}{kT}\right),$$

где n_i и W_{Fi} — равновесная концентрация и уровень Ферми в собственном полупроводнике, откуда следует важное соотношение $p_p n_p = n_i^2$.

Это — закон действующих масс для полупроводника p -типа. Он справедлив в случае термодинамического равновесия. Данный закон аналогичен закону действующих масс для полупроводника n -типа.

Из закона действующих масс следует, что при комнатной температуре концентрация неосновных носителей очень мала. Так, если $N_A = 1 \cdot 10^{16} \text{ см}^{-3}$, то при комнатной температуре в кремнии концентрация электронов определяется по формуле

$$n_p = \frac{n_i^2}{p_p} = \frac{n_i^2}{N_A} = \frac{(1,4 \cdot 10^{10})^2}{(1 \cdot 10^{16})} \approx 2 \cdot 10^4 \text{ см}^{-3}$$

и составляет всего $20\,000 \text{ см}^{-3}$, т.е. на 12 порядков меньше, чем концентрация дырок $p_p \gg n_p$. А для германия, при той же степени легирования и температуре, концентрация электронов составит

$$n_p = \frac{n_i^2}{N_A} = \frac{(2,5 \cdot 10^{13})^2}{(1 \cdot 10^{16})} \approx 6,25 \cdot 10^{10} \text{ см}^{-3},$$

т.е. электронов (неосновных носителей) на 6 порядков меньше, чем дырок.

В германии неосновных носителей заряда будет на 6 порядков больше, чем в кремнии.

Рассмотрим зависимость концентрации носителей от температуры (рис. 16).

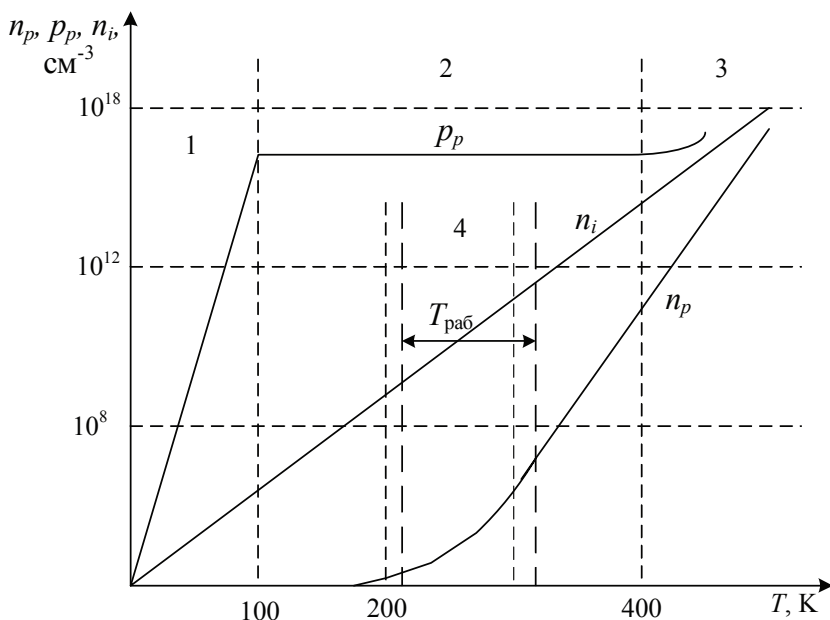


Рис. 16. Зависимость концентрации носителей от температуры

В собственном полупроводнике n_i и p_i возрастают с ростом температуры по экспоненциальному закону (в таком масштабе осей это будет прямая линия)

$$n_i \sim \exp\left(-\frac{\Delta W_3}{2kT}\right).$$

Концентрация основных носителей заряда — дырок — изменяется более сложным образом. В области очень низких температур при увеличении температуры происходит увеличение p_p за счет ионизации акцепторных атомов (участок 1 на рис. 16). В рабочем интервале температур (примерно $-100\dots+100^\circ\text{C}$) концентрация p_p сохраняется приблизительно постоянной и равной концентрации примеси, т. к. все акцепторные атомы ионизированы, а процесс тепловой генерации (ионизация атомов собственного вещества) добавляет относительно небольшое число основных носителей заряда (участок 2 на рис. 16). Однако концентрация неосновных носителей заряда, несмотря на их малость, возрастает очень быстро, и, при некоторой критической температуре, полупроводник теряет примесный характер проводимости (участок 3 на рис. 16), т. е. $n_i > N_A$, $p_p > N_A$ и $n_p > N_A$.

Из вышесказанного можно сделать следующие выводы:

- 1) концентрация как основных, так и неосновных носителей заряда зависит от положения уровня Ферми;
- 2) введение в полупроводник примесей сдвигает уровень Ферми относительно середины запрещенной зоны в дырочном полупроводнике вниз к валентной зоне;
- 3) повышение температуры полупроводника сдвигает уровень Ферми к середине запрещенной зоны;
- 4) увеличение концентрации примесей увеличивает концентрацию основных носителей заряда — дырок и уменьшает концентрацию неосновных носителей заряда — электронов;
- 5) при равной степени легирования, неосновных носителей заряда существенно меньше в полупроводниках с большей запрещенной зоной, а значит, с меньшей собственной концентрацией носителей (например, в кремнии на 6 порядков меньше неосновных носителей, чем в германии);
- 6) для увеличения диапазона рабочих температур $T_{\text{раб}}$ необходимо —
 - использовать полупроводниковый материал с большей запрещенной зоной;
 - увеличить степень легирования полупроводника.

3. УРОВЕНЬ ФЕРМИ

Для определения числа частиц, имеющих энергию в заданном интервале, помимо плотности квантовых состояний $N(W)$ необходимо знать вероятность того, что данное состояние с энергией W занято частицей, т.е. нужно знать функцию распределения $f(W)$. В условиях теплового равновесия для частиц с полуцелым спином, подчиняющихся принципу Паули, справедливо распределение Ферми — Дирака

$$f_n(W) = \frac{1}{e^{(W-W_F)/kT} + 1},$$

где W_F — энергия Ферми, или электрохимический потенциал — работа, которую необходимо затратить для изменения числа частиц в системе на единицу при условии постоянства объема и температуры; k — постоянная Больцмана; T — абсолютная температура.

Рассмотрим вид функции распределения Ферми — Дирака при различных температурах. Из вышеприведенной формулы следует, что в случае $T = 0$ в интервале энергии $0 \leq W < W_F$ имеем $f_n = 1$ и $f_n = 0$ для $W > W_F$. Это означает, что все квантовые состояния с энергией меньшей энергии Ферми заняты электронами, а уровни, лежащие выше уровня Ферми, полностью свободны, не заняты электронами. Следовательно, энергия Ферми есть максимально возможная энергия электронов в металле при температуре абсолютного нуля.

Рассмотрим случай, когда $T > 0$. Из распределения Ферми — Дирака для значения энергии, равного значению энергии Ферми ($W = W_F$), имеем $f_n = 1/2$. Таким образом, уровень Ферми есть энергетический уровень, вероятность заполнения которого при температуре, отличной от абсолютного нуля, равна 0,5. При $T > 0$ часть электронов в результате теплового движения перейдет в состояния с энергией, большей энергии ферми ($W > W_F$), и соответственно часть состояний, находящихся ниже уровня Ферми, окажется свободной. В этом случае

число частиц, перешедших на более высокие энергетические уровни, будет равно количеству образовавшихся свободных состояний в области $W < W_F$.

Произведем оценку области изменения функции распределения $f_n(W)$ для случая $T > 0$. Для этого подсчитаем $f_n(W)$ для разных значений энергии. Для энергий, отличающихся от W_F на $(\pm kT)$, значение на $f_n(W)$ составляет $(1 + e)^{-1} = 0,27$ и $(1 + 1/e)^{-1} = 0,73$. При $W - W_F = (\pm 2kT)$ значения f_n равны 0,118 и 0,882, а при $W - W_F = (\pm 3kT)$ — 0,047 и 0,953. Из этих данных следует, что вероятность заполнения состояний заметно отличается от единицы или нуля лишь в пределах $(2-3) kT$ вблизи значения $W = W_F$ (рис. 17).

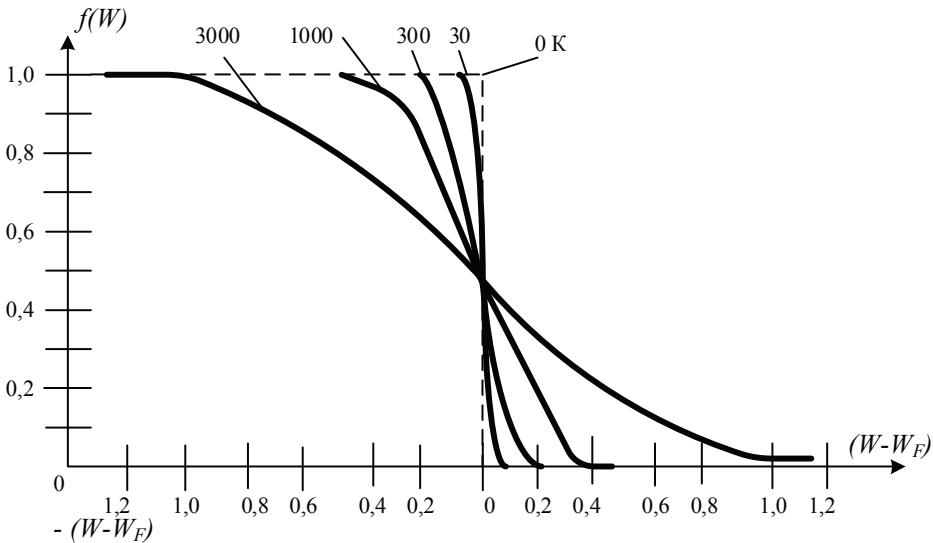


Рис. 17. Вид функции распределения Ферми — Дирака

Функция распределения Ферми — Дирака характеризует вероятность заполнения данного квантового состояния электроном. Вероятность того, что при тепловом равновесии в состоянии с энергией W электрон отсутствует, т. е. оно занято дыркой, составит

$$f_p(W) = 1 - f_n(W) = 1 - \frac{1}{e^{(W - W_F)/kT} + 1} = \frac{1}{e^{(W_F - W)/kT} + 1}.$$

Следовательно, функция распределения для дырок аналогична функции распределения для электронов, если отсчитывать энергию

дырок от уровня Ферми в противоположную сторону по сравнению с направлением отсчета энергии для электронов.

Для электронов, находящихся в состояниях с энергией $W - W_F \gg kT$, выражения для f_n и f_p имеют вид

$$f_n = \exp\left(-\frac{W - W_F}{kT}\right),$$

$$f_p = \exp\left(-\frac{W_F - W}{kT}\right),$$

т.е. совпадают с функцией распределения Больцмана для частиц, подчиняющихся классической статистике. Если носители заряда подчиняются статистике Больцмана, то электронный газ невырожден, полупроводник с таким распределением носителей заряда принято называть невырожденным.

Таким образом, для большинства полупроводников (невырожденных) можно пользоваться статистикой Максвелла — Больцмана и только в некоторых случаях для полупроводников (вырожденных) необходимо использовать статистику Ферми — Дирака. Разница в этих двух функциях распределения электронов по энергиям показана на рис. 18.

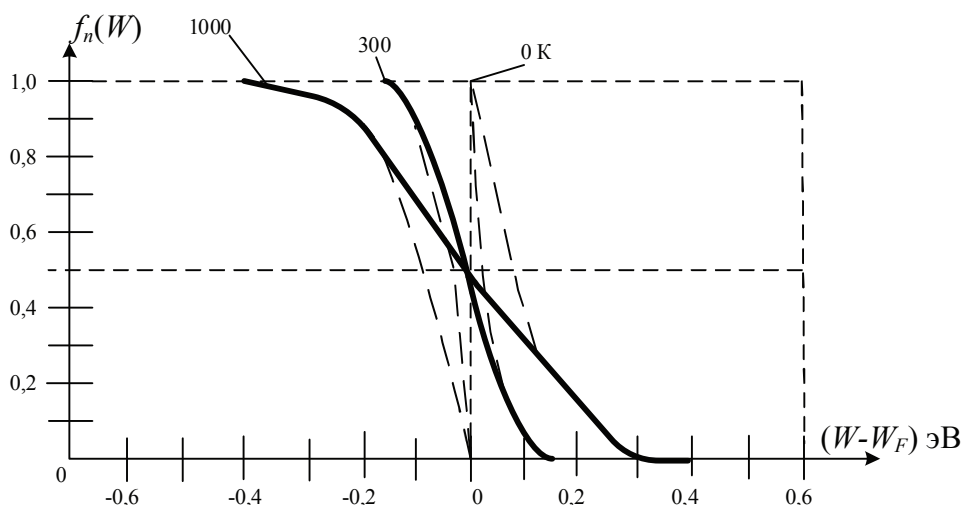


Рис. 18. Вероятность заполнения электронами энергетических уровней при различных температурах:

сплошная линия — по статистике Ферми — Дирака; пунктир — по статистике Максвелла — Больцмана для электронов в зоне проводимости и валентной зоне

Положение уровня Ферми в полупроводнике будет определять дрейфовую, и диффузионную составляющие тока.

Одно из фундаментальных положений физики твердого тела — постоянство (одинаковость) уровня Ферми для всех частей равновесной системы твердых тел, какой бы разнородной она не была. Другими словами, в условиях равновесия, когда направленного движения носителей заряда нет, должно иметь место условие $W_F = \text{const}$, т. е. $\text{grad } W_F = 0$, тогда ток в полупроводнике $I = 0$.

Для собственного полупроводника уровень Ферми определяется по выражению

$$W_{Fi} = \frac{W_{\Pi} + W_{\text{В}}}{2} + \frac{3}{4} kT \ln \frac{m_p^*}{m_n^*},$$

где W_{Π} — энергетический уровень дна зоны проводимости; $W_{\text{В}}$ — энергетический уровень потолка валентной зоны; m_p^*, m_n^* — эффективная масса дырок и электронов соответственно.

При температуре абсолютного нуля, уровень Ферми для собственного полупроводника лежит в середине запрещенной зоны. У собственного полупроводника скорость изменения положения уровня Ферми с температурой пропорциональна отношению эффективных масс дырок и электронов. В результате этого, с повышением температуры, уровень Ферми отдалется от зоны с тяжелыми носителями заряда, приближаясь к зоне с легкими носителями заряда. Например, при $m_p^* > m_n^*$ уровень Ферми с повышением температуры линейно смещается к дну зоны проводимости. И если расстояние от уровня Ферми до этой зоны становится соизмеримо с kT , то в ней наступает вырождение и соответствующий интеграл Ферми — Дирака уже не может быть заменен экспонентой. При этом чем сильнее различаются эффективные массы электронов и дырок, тем раньше наступает вырождение.

В случае $m_p^* = m_n^*$ положение уровня Ферми не зависит от температуры и определяется по положению середины запрещенной зоны

$$W_{Fi} = \frac{W_{\Pi} + W_{\text{В}}}{2}.$$

Ширина запрещенной зоны изменяется с температурой. Рост амплитуды тепловых колебаний атомов решетки приводит к ее уменьшению. Кроме того, с увеличением температуры изменяются межатомные расстояния, что также оказывает влияние на ширину запрещенной зоны. Зависимость $\Delta W_3(T)$ может иметь сложный характер. В каче-

стве примера показаны изменения ширины запрещенной зоны в зависимости от температуры для германия, кремния и арсенида галлия (рис. 19). Для этих полупроводников значения ширины запрещенной зоны при 0 К составляют 0,89; 1,16 и 1,52 эВ соответственно. У них, как следует из рис. 19, в диапазоне температур 175–350 К ширина запрещенной зоны меняется линейно с изменением температуры. В таком случае температурный коэффициент изменения ширины запрещенной зоны $\alpha = d\Delta W_3/dT < 0$ зависит от материала полупроводника: Ge = $3,9 \cdot 10^{-4}$ эВ/К, Si = $2,4 \cdot 10^{-4}$ эВ/К, GaAs = $4,3 \cdot 10^{-4}$ эВ/К, InSb — до 1750 эВ/К. У PbS $\alpha < 0$, ширина запрещенной зоны возрастает от 0,34 эВ при 0 К до 0,41 эВ при 300 К.

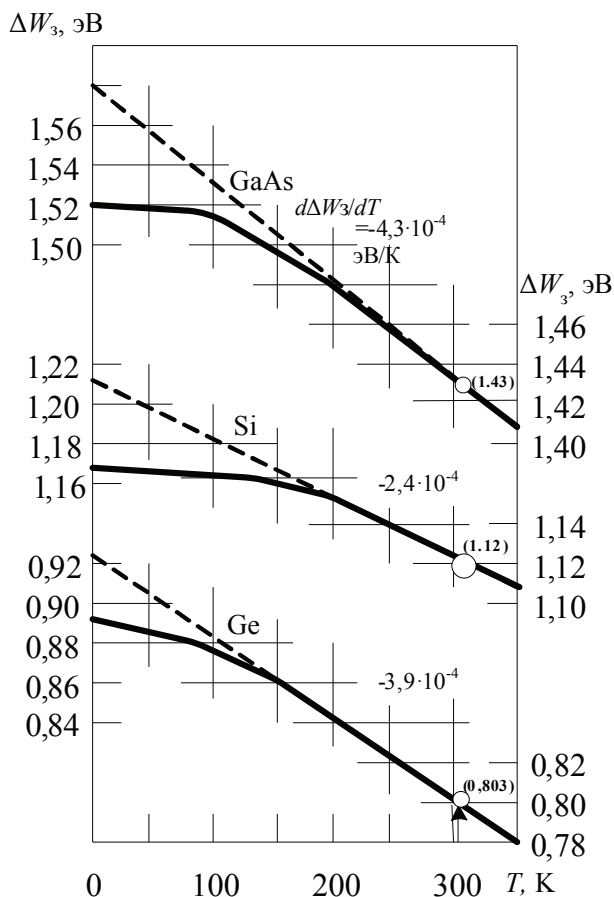


Рис. 19. Зависимость ширины запрещенной зоны германия, кремния и арсенида галлия от температуры

В этих случаях зависимость подчиняется линейному закону вида

$$\Delta W_3 = \Delta W_3(0) - \alpha T,$$

где $\Delta W_3(0)$ — экстраполированная ширина запрещенной зоны при 0 К.

Теоретический анализ показывает следующую зависимость собственной концентрации электронов:

$$\ln(n_i T^{-3/2}) = \text{const} - \frac{\alpha}{2k} - \frac{\Delta W_0}{2k} \frac{1}{T}.$$

откуда следует, что

$$\ln(n_i T^{-3/2}) = \text{const} - \alpha / 2k - (\Delta W_0) / 2kT.$$

Если ширина запрещенной зоны полупроводника линейно зависит от температуры, график зависимости $\ln(n_i T^{3/2})$ от $1/T$ также представляет собой прямую линию, наклон которой характеризуется значением $\Delta W_3(0)$, которое является экстраполированной шириной запрещенной зоны при 0 К. Истинное значение ширины запрещенной зоны полупроводника при данной температуре определяется по формуле $\Delta W_3 = \Delta W_3(0) - \alpha T$.

Для примесных полупроводников уровень Ферми можно определить из соотношений (справедливы для $T \neq 0$ К):

$$W_{Fn} = W_{\Pi} - kT \ln \frac{N_C}{N_D},$$

$$W_{Fp} = W_{\text{В}} - kT \ln \frac{N_V}{N_A},$$

где N_C, N_V — эффективная плотность разрешенных уровней в зоне проводимости и валентной зоне соответственно; N_D, N_A — количество донорных и акцепторных уровней (степень легирования).

При решении задач удобнее использовать соотношения для вычисления положения уровня Ферми относительно середины запрещенной зоны (справедливы для $T \neq 0$ К):

$$W_{Fn} = \frac{1}{2}(W_{\Pi} + W_{\text{В}}) + kT \ln \frac{n_n}{n_i},$$

$$W_{Fp} = \frac{1}{2}(W_{\Pi} + W_{\text{В}}) - kT \ln \frac{p_p}{n_i}.$$

Таким образом, положение уровня Ферми в примесных полупроводниках зависит от температуры, степени легирования и ширины запрещенной зоны.

Для определения поведения уровня Ферми в области низких температур необходимо уточнить функцию Ферми — Дирака для примесных полупроводников.

Рассмотрим полупроводник, содержащий донорную примесь с концентрацией N_D . Если бы на примесном уровне согласно принципу Паули могли расположиться 2 электрона с антипараллельными спинами, то вероятность его заполнения определялась бы функцией Ферми — Дирака

$$f_n(W) = \frac{1}{e^{(W - W_F)/kT} + 1},$$

в которой вместо W следовало бы поставить W_D — энергию электрона на уровне примеси. Но на уровне W_D может быть только один электрон (атом донора может удерживать один электрон). Тогда вероятность нахождения электрона на донорном уровне с энергией W_D будет определяться по выражению

$$f_n(W_D) = \frac{1}{\frac{1}{2} e^{(W_D - W_F)/kT} + 1}.$$

Предэкспоненциальный множитель $1/2$ можно записать через g^{-1} . И для одновалентной донорной примеси (может отдать для участия в проводимости только 1 электрон), примесный уровень двукратно вырожден и фактор (степень) спинового вырождения $g = 2$.

Аналогично полупроводнику с донорными примесями, для акцепторного полупроводника, например кремния, легированного бором, будет определяться вероятность нахождения электрона на акцепторном уровне. Нейтральный атом бора с соседними атомами кремния образует 3 ковалентных связи, четвертая связь одного из четырех соседних атомов кремния остается незавершенной, и она, располагаясь около атома бора, ведет себя как положительная дырка. В эту незавершенную связь может перейти электрон от соседнего атома кремния, для этого потребуется энергия равная W_A . В результате этого образуется свободная дырка, а атом бора превращается в отрицательно заряженный ион бора. Таким образом, на энергетическом уровне акцепторной примеси находится 1 электрон с произвольным направле-

нием спина (нейтральное состояние акцепторной примеси) либо имеется 2 электрона с антипараллельными спинами в случае, когда атом акцепторной примеси для укомплектования парной связи захватывает электрон из валентной зоны (ионизированное состояние акцепторной примеси). Следовательно, степень вырождения акцепторного уровня $g = 2$.

В области низких температур (рис. 20) положение уровня Ферми будет определяться по соотношению вида

$$W_F = W_{\Pi} - \frac{W_{\Pi} - W_{\text{д}}}{2} - \frac{kT}{2} \ln \frac{gN_c}{N_{\text{д}}} = \frac{W_{\Pi} + W_{\text{д}}}{2} - \frac{kT}{2} \ln \frac{gN_c}{N_{\text{д}}},$$

где g — фактор спинового вырождения,
а энергия активации будет

$$W_{\text{ион}} = \frac{1}{2}(W_{\Pi} - W_{\text{д}}),$$

т. е. равна половине энергии ионизации донорной примеси. В невырожденном донорном полупроводнике, при температуре абсолютно нуля, уровень Ферми располагается посередине между дном зоны проводимости и уровнем донорной примеси.

Строгий теоретический анализ показывает, что в области достаточно низких температур (несколько градусов по шкале Кельвина), когда $gN_c < N_{\text{д}}$, уровень Ферми сначала повышается до некоторого максимального значения, а затем начинает снижаться и при $gN_c = N_{\text{д}}$ снова $W_F = 1/2 (W_{\Pi} + W_{\text{д}})$, как и в случае $T = 0$. Дальнейшее повышение температуры сопровождается ростом N_c , и в области температур, когда $gN_c > N_{\text{д}}$, уровень Ферми продолжает снижаться. Такому перемещению уровня Ферми соответствует экспоненциальная температурная зависимость концентрации электронов

$$n_0 = N_c e^{\eta} = N_c e^{-(W_{\Pi} - W_{\text{д}})/2kT - \frac{1}{2} \ln(gN_c - N_{\text{д}})} = (g^{-1} N_c N_{\text{д}})^{1/2} e^{-(W_{\Pi} - W_{\text{д}})/2kT}.$$

Эта область изменения уровня Ферми является областью слабой ионизации примеси (или областью вымораживания) (1 на рис. 20).

При дальнейшем повышении температуры, концентрация электронов в зоне проводимости становится сравнимой с концентрацией примеси, и предыдущие выражения для W_F и n_n в этом случае неприменимы. Однако теперь можно рассматривать другой случай, когда температура достаточно высока, в таком случае выполняется неравенство

$$e^{(W_{\Pi}-W_{\text{Д}})/kT} \ll N_c / 2N_{\text{Д}},$$

где функция Ферми аппроксимируется по выражению

$$W_F = W_{\Pi} - kT \ln(N_c / N_{\text{Д}}),$$

которому соответствует $n_n = N_{\text{Д}}$.

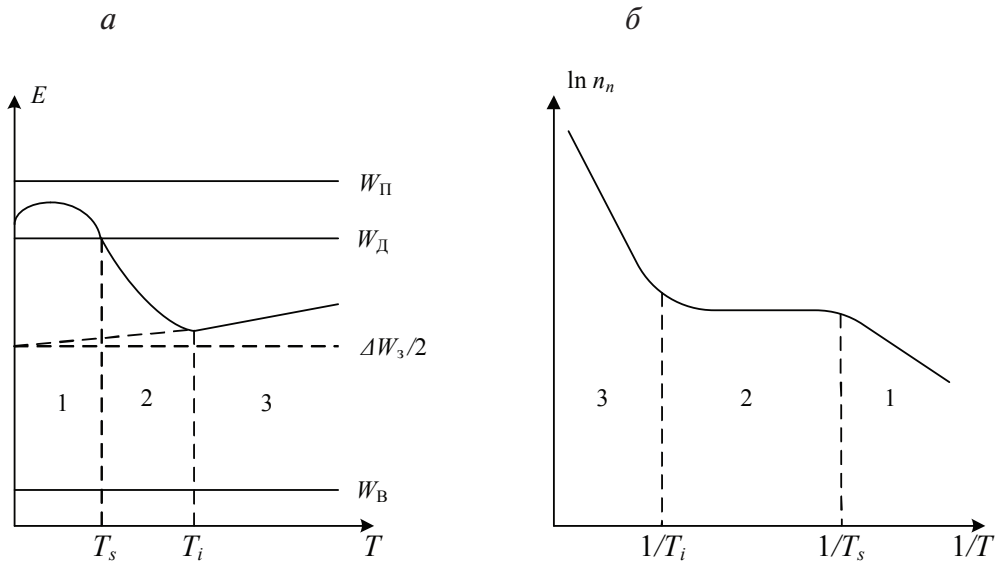


Рис. 20. Изменение положения уровня Ферми (а) и концентрации электронов (б) в зависимости от температуры для донорного полупроводника

Равенство $n_n = N_{\text{Д}}$ означает, что практически вся донорная примесь ионизирована, концентрация электронов в зоне проводимости не зависит от температуры. Данная область температур, при которой имеет место полная ионизация примеси, носит название области истощения примеси (или области — полной ионизации примеси) (на рис. 20 отмечена цифрой 2).

Условие полной ионизации донорной примеси, когда $n_n = N_{\text{Д}}$, соответствует положению уровня Ферми на несколько kT ниже уровня примеси $W_{\text{Д}}$. Это значит, что при повышении температуры, уровень Ферми, понижаясь, пересекает уровень $W_{\text{Д}}$ и уходит вниз. Темпера-

тура, при которой $W_F = W_D$, носит название температуры истощения T_s , ее можно определить из условия

$$W_F = W_D - kT_s \ln(N_c / N_D) = W_D,$$

откуда

$$T_s = \frac{W_D - W_F}{k \ln(N_c / N_D)}.$$

Как следует из данного выражения, температура истощения тем ниже, чем меньше энергия ионизации ($W_D - W_F$) и концентрация донорной примеси N_D и чем больше эффективная масса электронов, определяющая величину N_c . При малых значениях ($W_D - W_F$) истощение примеси наступает при очень низких температурах. Например, в электронном германии, легированном сурьмой в количестве $N_D = 10^{16} \text{ см}^{-3}$, для которой энергия ионизации равна 0,0096 эВ, насыщение наступает уже при $T_s = 32 \text{ К}$.

При дальнейшей повышении температуры, увеличение концентрации электронов в зоне проводимости будет происходить за счет переходов электронов из валентной зоны. В этом случае положение уровня Ферми и концентрация электронов будут определяться по уравнениям для W_{Fi} и n_i . На рис. 20 область 3 соответствует области собственной проводимости, тогда W_{Fi} и можно определить

$$W_{Fi} = W_D - \frac{1}{2} \cdot \Delta W_3 - \frac{kT_i}{2} \ln \frac{N_c}{N_v} = W_D - kT_i \ln \frac{N_D}{N_c}.$$

Отсюда получаем

$$T_i = \frac{\Delta W_3}{k \ln \left(\frac{N_c N_v}{N_D^2} \right)}.$$

Анализ этого выражения показывает, что температура T_i , при которой наступает собственная проводимость у донорного полупроводника, тем ниже, чем меньше ширина запрещенной зоны и концентрация примеси и чем больше значение эффективных масс носителей заряда.

Таким образом, используя описанные приближения, можно проследить изменение концентрации электронов и положения уровня Ферми в запрещенной зоне электронного полупроводника во всей

области изменения температуры. В качестве примера на рис. 21 приведены температурные зависимости уровня Ферми и концентрации равновесных электронов n_0 и дырок p_0 для германия, легированного сурьмой в количестве $N_d \approx 10^{16} \text{ см}^{-3}$. Кроме того, на этих кривых пунктиром показан ход W_{Fi} и n_i в собственном германии. При построении графиков учтена зависимость ширины запрещенной зоны германия от температуры.

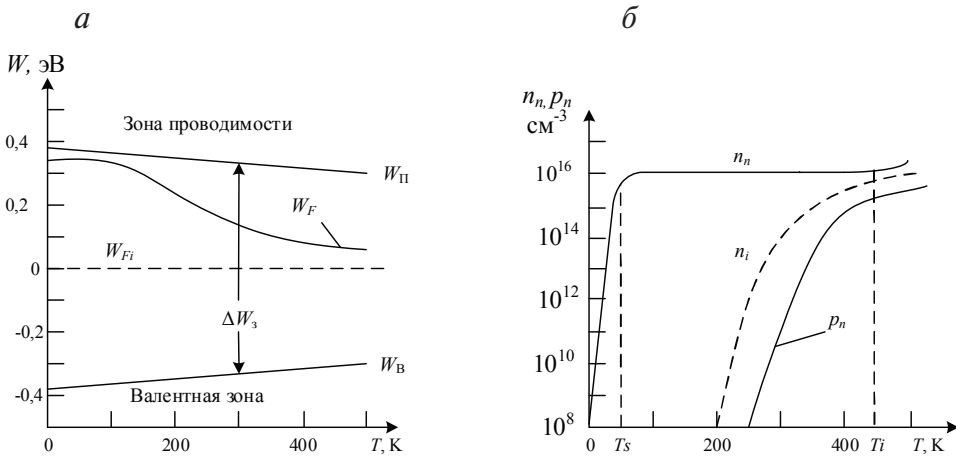


Рис. 21. Температурная зависимость уровня Ферми (а) и концентрации носителей заряда (б) для германия, легированного сурьмой

При температуре абсолютного нуля, уровень Ферми в германии расположен посередине между дном зоны проводимости W_{Π} и уровнем донорной примеси W_d (см. рис. 21). При повышении температуры он опускается и приближается к уровню примеси W_d . При температуре насыщения T_s на донорной примеси электроны находятся в количестве

$$n_n = \frac{N_d}{\frac{1}{2} e^{\frac{W_d - W_F}{kT}} + 1} = \frac{2}{3} N_d,$$

а в зоне проводимости равны соответственно $1/3 N_d$ электронов. С дальнейшим ростом температуры, уровень Ферми продолжает опускаться и наступает область истощения; вся примесь ионизирована, и концентрация электронов проводимости остается постоянной $n_n = N_d$. В этой температурной области имеет место уже ионизация атомов основного вещества и появляются неосновные носители за-

ряда — дырки. Их концентрация резко возрастает с ростом температуры согласно соотношению

$$p_n = \frac{n_i^2}{n_n} = \frac{N_c N_v}{N_d} e^{-\Delta W_3 / kT}.$$

Когда уровень Ферми достигает середины запрещенной зоны, тогда $n_n = p_n = n_i$ и полупроводник от примесного переходит к собственному. При дальнейшем повышении температуры уровень Ферми приближается к той зоне, которая имеет меньшую эффективную плотность состояний (рис. 22).

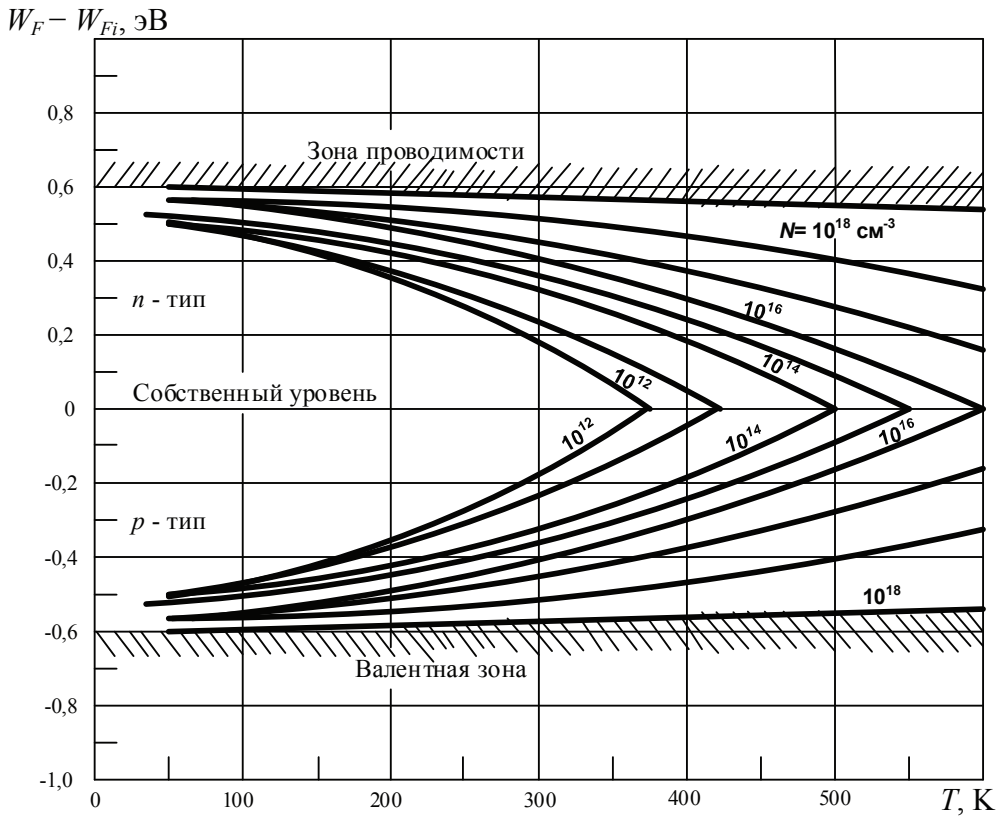


Рис. 22. Зависимость уровня Ферми для кремния от температуры и концентрации примесей

В акцепторном полупроводнике, как и в случае донорной примеси, при высоких температурах наступает область истощения, характери-

зующаяся полной ионизацией атомов акцепторной примеси. С дальнейшим ростом температуры, уровень Ферми поднимается к середине запрещенной зоны и полупроводник ведет себя как собственный.

Выводы:

- 1) при $T = 300$ °К уровень Ферми в n -полупроводнике лежит, как правило, ниже уровня донорной примеси W_D , но выше середины запрещенной зоны;
- 2) в p -полупроводнике уровень Ферми расположен выше уровня акцепторной примеси W_A , но ниже середины запрещенной зоны;
- 3) чем сильнее легирован полупроводник n -типа, тем ближе уровень Ферми ко дну зоны проводимости; для p -типа чем больше акцепторной примеси, тем ближе уровень Ферми к потолку валентной зоны. Таким образом, чем сильнее легирован полупроводник, тем ближе уровень Ферми к зоне, отвечающей за тип проводимости (зона основных носителей заряда);
- 4) с ростом температуры уровень Ферми в n -полупроводнике снижается к середине запрещенной зоны, а в p -полупроводнике повышается к середине запрещенной зоны, т.е. примесный полупроводник ведет себя как собственный;
- 5) чем сильнее легирован материал, тем выше максимальная рабочая температура прибора, использующего примесный характер полупроводника.

Задачи к ПРАКТИЧЕСКОМУ ЗАНЯТИЮ № 1

- 1) Найти положение уровня Ферми в собственном германии при 300 К, если известно, что эффективные массы плотности состояний для дырок валентной зоны и для электронов зоны проводимости соответственно равны $m_v = 0,388m_0$; $m_c = 0,55m_0$, где m_0 — масса свободного электрона.
- 2) Определить, на сколько различаются вероятности заполнения электронами нижнего уровня зоны проводимости в собственном германии и собственном кремнии: а) при 300 К; б) при 100 К.
- 3) Вычислить положение уровня Ферми при $T = 300$ К в кристаллах германия, содержащих $2 \cdot 10^{22} \text{ м}^{-3}$ атомов мышьяка и 10^{22} м^{-3} атомов галлия.

- 4) Вычислить, во сколько раз различаются равновесные концентрации дырок при комнатной температуре в кристаллах кремния и арсенида галлия, имеющих одинаковую концентрацию донорных примесей $N_d = 10^{21} \text{ м}^{-3}$. Для расчета использовать данные: $T = 300 \text{ К}$, эффективные массы плотности состояний $m_v = 0,56m_0$, $m_c = 1,05m_0$.
- 5) Удельное сопротивление собственного германия при $T = 300 \text{ К}$ $\rho = 0,44 \text{ Ом}\cdot\text{м}$. Определите собственную концентрацию электронов и дырок.

Задания: решение задач 4–6 раздела «Однородный полупроводник» (приложение «Домашние задания»).

4. ДРЕЙФ. ЭЛЕКТРОПРОВОДНОСТЬ

В полупроводниках свободные электроны и дырки находятся в состоянии хаотического движения. Поэтому, если выбрать произвольное сечение внутри объема полупроводника и подсчитать число носителей заряда, проходящих через это сечение за единицу времени слева направо и справа налево, значения этих чисел окажутся одинаковыми. Это означает, что электрический ток в данном объеме полупроводника отсутствует.

При помещении полупроводника в электрическое поле напряженностью E , на хаотическое движение носителей зарядов накладывается составляющая направленного движения, т. е. под действием разности потенциалов в полупроводнике возникает электрическое поле, которое ускоряет электроны и дырки и сообщает им еще некоторое поступательное движение, представляющее собой ток проводимости. Движение носителей заряда под действием электрического поля называют дрейфом носителей, а ток проводимости — дрейфовым током I_E . Полный ток проводимости складывается из электронного I_{En} и дырочно-го I_{Ep} тока проводимости

$$I_E = I_{En} + I_{Ep}.$$

Несмотря на то что электроны и дырки движутся в противоположных направлениях, эти токи складываются (рис. 23), т. к. движение дырок представляет собой перемещение электронов.

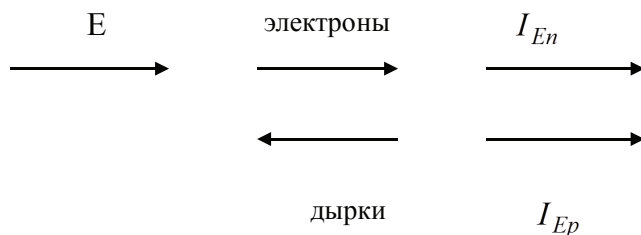


Рис. 23. Дрейфовый ток в полупроводнике

Следует обратить внимание на направления вектора напряженности поля и движения носителей заряда, а также на связь вектора напряженности с вектором дрейфового тока.

Из-за столкновения носителей зарядов с атомами кристаллической решетки, их движение в направлении действия электрического поля прерывисто. На рис. 24 показаны процессы рассеивания и дрейфа электрона в полупроводниковом кристалле, которые начались в точке 0. В черных точках электрон рассеивается и меняет вектор скорости. Рассеивание носит случайный характер.

Чтобы установить, от каких величин зависит ток дрейфа, удобнее рассматривать не сам ток, а его плотность.

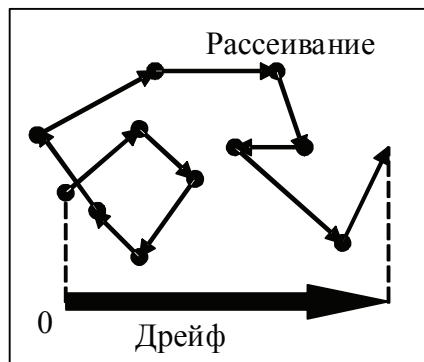


Рис. 24. Процессы рассеяния и дрейфа носителей

Очевидно, что плотность тока дрейфа I_E складывается из плотностей электронного и дырочного токов

$$i_E = i_{En} + i_{Ep}.$$

Поскольку плотность тока равна количеству электричества, проходящему через единицу площади поперечного сечения за 1 с, плотность электронного тока составит

$$i_{En} = qnv_{др},$$

где q — заряд электрона; n — концентрация электронов и $v_{др}$ — средняя скорость дрейфа.

Скоростью дрейфа называется скорость, направленная вдоль вектора напряженности электрического поля, усредненная по всем носителям одного знака (электронам или дыркам). Если движение носителей в промежутке между двумя последовательными взаимодействиями

с рассеивающими центрами является равноускоренным, то скорость дрейфа будет

$$\overline{v_{\text{др}}} = \overline{t_{\text{П}}} q E / m = \mu E ,$$

где $\overline{v_{\text{др}}}$ — скорость дрейфа; q — заряд электрона; E — напряженность электрического поля; m — эффективная масса свободных носителей; μ — подвижность электронов.

Коэффициент пропорциональности μ , между скоростью дрейфа и напряженностью электрического поля, является основным параметром дрейфового движения и называется подвижностью. Подвижность характеризует скорость дрейфа, приобретаемую свободными носителями в электрическом поле единичной напряженности, например 1 В/см, в таком случае $\mu = 1 \text{ см}^2/(\text{В} \cdot \text{с})$. Выразив скорость дрейфа через $\mu_n E$, получим

$$i_{En} = q n \mu E ,$$

где произведение $q n \mu$ представляет собой удельную электронную проводимость σ_n . Плотность дрейфового тока тогда запишем в виде

$$i_{En} = \sigma_n E .$$

Удельная проводимость зависит от концентрации носителей n и от их подвижности μ .

Что же такое подвижность?

В соответствии с выражением для $\overline{v_{\text{др}}}$ имеем $\mu = q \overline{t_{\text{П}}} / m$. Здесь среднее время свободного пробега $\overline{t_{\text{П}}}$ можно выразить через среднюю длину свободного пробега \bar{l} и среднюю скорость $\bar{v} = q \bar{l} / (m \bar{v})$.

Среднее время свободного пробега $\overline{t_{\text{П}}}$ и средняя длина свободного пробега \bar{l} характеризуют частоту столкновения носителей заряда с теми или иными «препятствиями». В результате таких столкновений происходит изменение скорости и направления движения носителей, т. е. их рассеяние. При температуре абсолютного нуля, в идеальной кристаллической решетке рассеяние не имеет места. Иначе говоря, атомы решетки как таковые не являются препятствиями на пути движения носителей. Истинными препятствиями являются лишь колеблющиеся атомы решетки — фононы, а также атомы примеси и дефекты структуры.

Подвижность носителей заряда зависит: от типа носителя заряда, материала полупроводника, температуры, концентрации примесей,

напряженности электрического поля, дефектов кристаллической решетки.

Рассмотрим влияние на подвижность материала полупроводника, типа носителя заряда, температуры, концентрации примесей и напряженности электрического поля.

1. Тип носителя.

Подвижность зависит от эффективной массы свободных носителей и имеет разные значения для электронов и дырок. Как правило, подвижность электронов больше подвижности дырок. Например, в германии $\mu_n > 2\mu_p$, а в арсениде галлия $\mu_n \gg \mu_p$.

2. Материал полупроводника.

Исследования показали, что подвижность электронов в различных полупроводниках различна, так, в германии подвижность любого носителя заряда выше, чем в кремнии $\mu_{\text{Ge}} > \mu_{\text{Si}}$.

Таблица 2

Подвижность носителей заряда в чистых полупроводниках при комнатной температуре

Подвижность $\text{см}^2/(\text{В}\cdot\text{с})$	Ge	Si	GaAs	SbIn
μ_n	3800	1400	11000	До 65000
μ_p	1800	500	450	700

Например, при комнатной температуре подвижность электронов в чистом германии составляет $3800 \text{ см}^2/(\text{В}\cdot\text{с})$, т. е. под действием поля напряженностью 1 В/см электроны проводимости чистого германия получают среднюю скорость 3800 см/с , а в чистом кремнии электроны проводимости получают среднюю скорость лишь 1400 см/с (табл. 2).

3. Температура.

Подвижность с ростом температуры будет падать

$$\mu = AT^{-3/2},$$

где A — коэффициент пропорциональности.

Падение подвижности носителей заряда с ростом температуры объясняется возрастанием числа их столкновений в единицу времени, т. е. сокращением времени свободного пробега. Другими словами, причина падения подвижности — рассеяние носителей заряда при тепловых колебаниях атомов кристаллической решетки. На практике зависи-

мость подвижности от температуры при решеточном рассеянии может отклоняться от этого закона. В общем случае $\mu \sim T^{-\alpha}$; например, для кремния *n*-типа $\alpha = 2,42$; для кремния *p*-типа $\alpha = 2,2$; для германия *n*- и *p*-типов — соответственно 1,66 и 2,33; для арсенида галлия — 1 и 2,1. Таким образом, при рассеянии свободных носителей при колебаниях решетки подвижность уменьшается с ростом температуры

$$\mu = \mu_0 \left(\frac{T}{T_0} \right)^{\alpha},$$

где μ_0 — подвижность при температуре T_0 , например, комнатной. Температурная зависимость подвижности (рис. 25) в примесном полупроводнике определяется решетчатым и ионным механизмами рассеяния.

4. Концентрации примесей.

Правильнее сказать, что на подвижность влияет не столько концентрация примесей, сколько электрические поля ионизированных примесей, т. е. происходит рассеяние на ионах примесей.

Поскольку ионы примесей являются центрами рассеивания в полупроводнике, то увеличение концентрации примесей приведет к уменьшению длины свободного пробега зарядов и к снижению подвижности. Зависимость подвижности от концентрации примесей показана на рис. 26.

Для достаточно больших концентраций ($N > N_0 = 10^{15} \dots 10^{16} \text{ см}^{-3}$)

$$\mu = \mu_0 \left(\frac{N_0}{N} \right)^{1/3},$$

где подвижность μ_0 соответствует концентрации N_0 .

Иногда удобнее использовать формулу

$$\mu = \mu_0 - \Delta\mu \lg \frac{N_0}{N},$$

в которой коэффициент $\Delta\mu$ есть изменение подвижности на декаду приращения концентрации.

При малой концентрации примесей преобладает рассеяние на фонах, подвижность имеет максимальное значение, соответствующее беспримесному полупроводнику. При больших концентрациях примесей, на подвижность влияет рассеяние на ионизированных атомах и подвижность уменьшается с ростом их концентрации. Таким образом, важнейшими факторами, определяющими значение подвижности, являются температура и концентрация примесей.

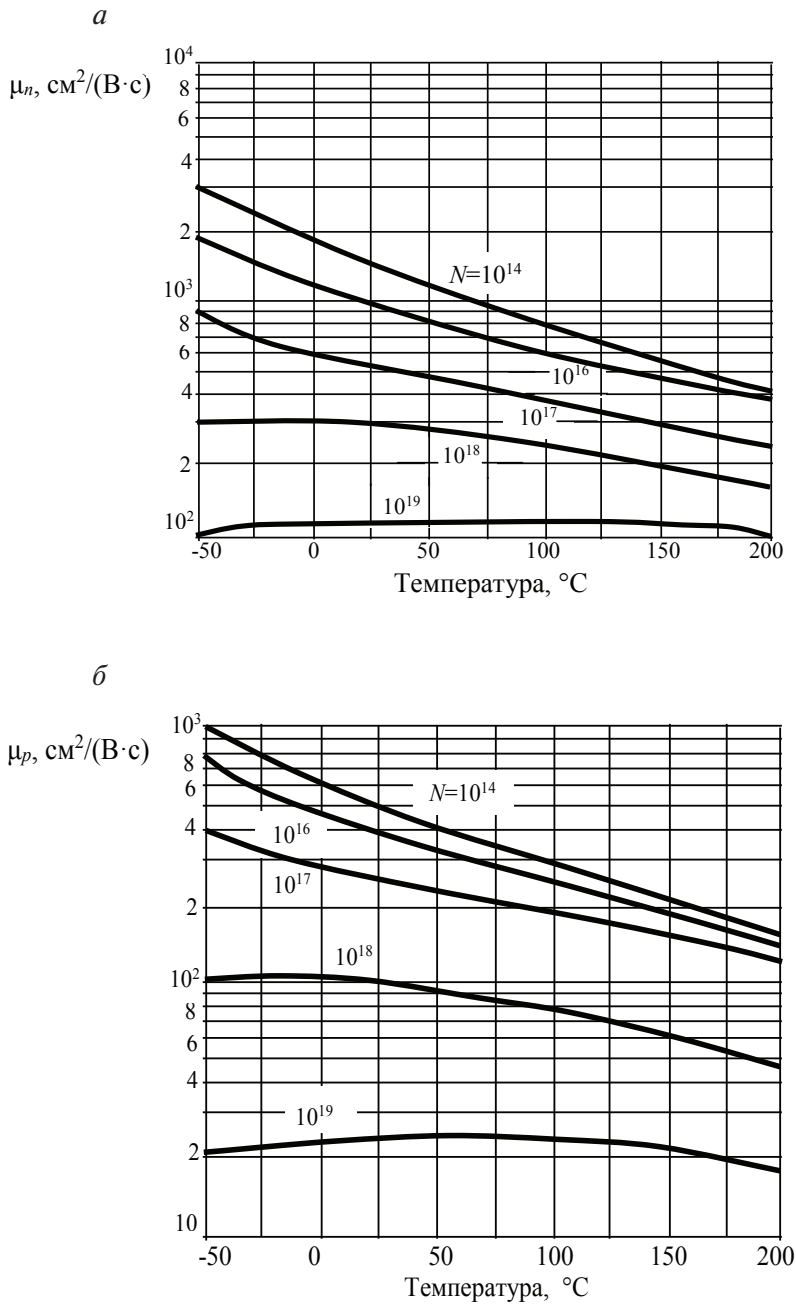


Рис. 25. Зависимости подвижности основных носителей от температуры при различных концентрациях примесей:

a — подвижность электронов в кремнии *n*-типа;
б — подвижность дырок в кремнии *p*-типа

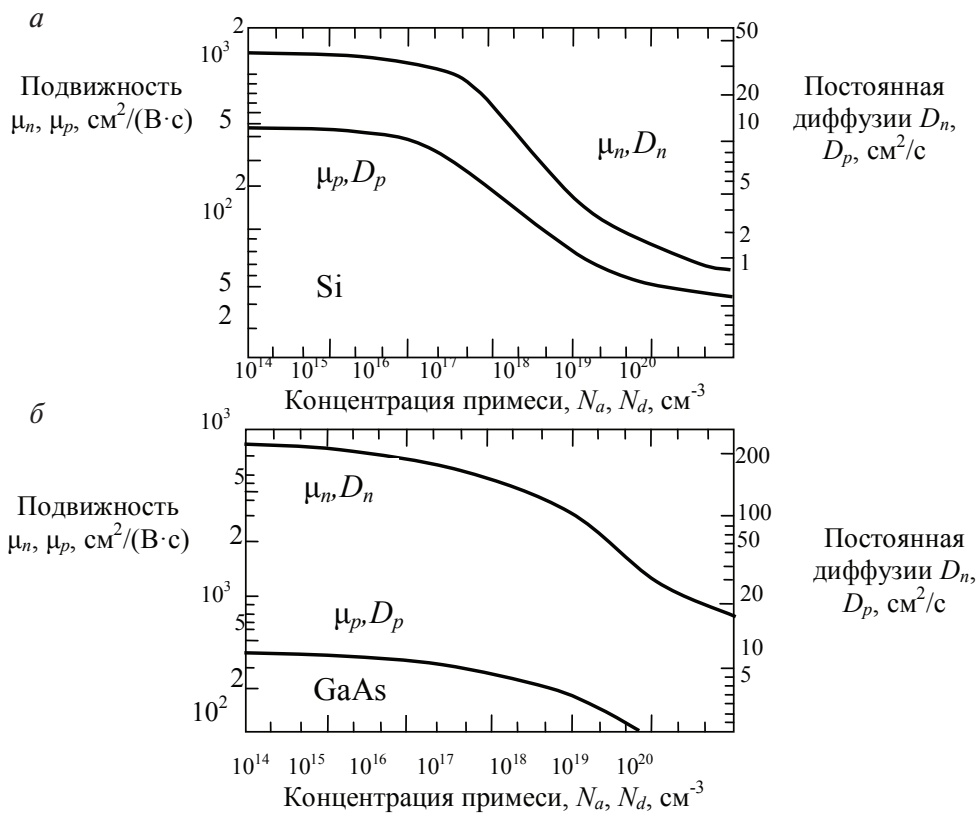


Рис. 26. Зависимости подвижности и коэффициента диффузии от концентрации примесей для полупроводников Si (а) и GaAs (б) при комнатной температуре

Иной характер изменения подвижности возникает в особом случае: при низких температурах и повышенных концентрациях примесей.

В таком случае основную роль играет рассеяние на ионизированных атомах примесей. Механизм рассеяния движущегося электрона неподвижным ионом показан на рис. 27.

Если температура низкая, то тепловая скорость электрона мала. В результате притяжения электрон «падает» на ион (траектория 1), превращая его на некоторое время в нейтральный атом, после чего, в результате теплового возбуждения, электрон отрывается от атома и начинает движение в случайном направлении, не связанном с первоначальным направлением. Это соответствует сильному рассеянию.

При большей температуре скорость электрона увеличивается. Он уже не «падает» на ион, хотя направление движения изменяется (траектория 2).

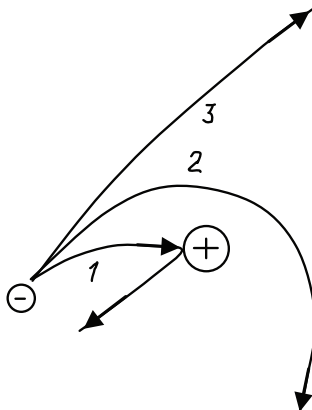


Рис. 27. Механизм рассеяния при низких температурах

При еще большей температуре, вследствие большой скорости, направление движения почти не изменяется (траектория 3), что соответствует слабому рассеянию. Таким образом, только при низких температурах при рассеянии свободных носителей на ионизированных примесях подвижность увеличивается с ростом температуры вследствие уменьшения времени их взаимодействия с ионами $\mu \sim T^{+3/2}$.

5. Напряженность электрического поля.

Зависимость дрейфовой скорости от напряженности электрического поля для германия, кремния и арсенида галлия при $T = 300$ К показана на рис. 28. Для германия и кремния эта зависимость может быть аппроксимирована по формуле $v_{др} = v_{нас} [1 + (E_0/E)]^{-1}$, где E_0 — параметр аппроксимации, $E_0 = v_{нас}/\mu_0$. В соответствии с этим зависимость подвижности от напряженности электрического поля может быть аппроксимирована по выражению

$$\mu = \mu_0 / [1 + (E\mu_0/v_{нас})].$$

Важную роль играет зависимость подвижности от напряженности электрического поля. Если напряженность превышает критическое значение $E_{кр}$, то подвижность

$$\mu = \mu_0 \left(\frac{E_{\text{кр}}}{E} \right)^{1/2},$$

где μ_0 — подвижность при $E = E_{\text{кр}}$, а $E > E_{\text{кр}}$.

На начальном участке соблюдается закон Ома, поскольку $\mu = \text{const}$, и, следовательно, $v \sim E$. На последующих двух участках этот закон все больше нарушается вплоть до появления участка насыщения: скорость дрейфа, а следовательно, и ток практически не зависят от напряженности электрического поля.

Режим $E > E_{\text{кр}}$ носит название сверхкритического режима, при котором происходит так называемый «разогрев» носителей.

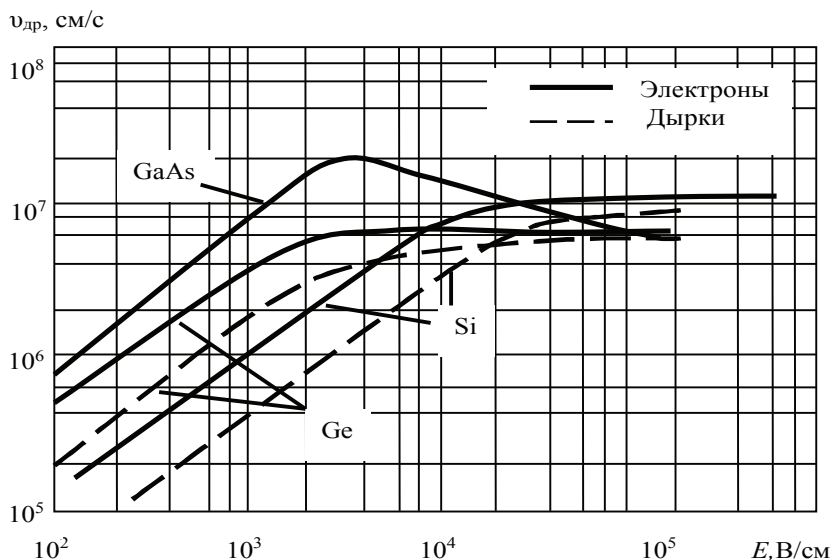


Рис. 28. Скорость дрейфа носителей в виде функции от напряженности электрического поля при комнатной температуре

Критическая напряженность поля не является строго определенной величиной; она лишь приблизительно соответствует условию, когда направленная (дрейфовая) скорость носителей делается сравнимой с их хаотической (тепловой) скоростью.

Ранее из закона Ома для плотности тока $i_{En} = \sigma_n E$ мы установили, что удельная проводимость зависит от концентрации носителей и от их

подвижности. Поскольку в полупроводнике имеется два типа носителей, удельная проводимость имеет две компоненты, т. е.

$$\sigma = \sigma_n + \sigma_p,$$

где σ_n — электронная, а σ_p — дырочная составляющие, $\sigma_n = qn\mu_n$; $\sigma_p = qp\mu_p$. Здесь q — заряд носителей; n и p — соответствующие концентрации носителей заряда; μ_n и μ_p — подвижности соответствующих носителей заряда.

Плотность полного тока дрейфа в собственном полупроводнике будет

$$i_{Ei} = i_{En} + i_{Ep} = qn_i\mu_n E + qp_i\mu_p E = qn_i(\mu_n + \mu_p)E,$$

а полная удельная проводимость собственного полупроводника —

$$\sigma_i = qn_i(\mu_n + \mu_p).$$

Получается, что удельная проводимость собственного полупроводника зависит от равновесной концентрации носителей n_i и подвижностей дырок и электронов. В полупроводниках, при повышении температуры, вследствие интенсивной генерации пар носителей, увеличение равновесной концентрации носителей заряда происходит значительно быстрее, нежели уменьшение их подвижности, поэтому с повышением температуры проводимость σ_i растет.

Напомним также, что $\mu_{Ge} > \mu_{Si}$, $n_{iGe} \gg n_{iSi}$ и, следовательно,

$$\sigma_{iGe} \gg \sigma_{iSi}.$$

Плотность полного тока дрейфа в полупроводнике n -типа определяется как

$$i_E = i_{En} + i_{Ep} = qn_n\mu_n E + qp_n\mu_p E \approx qn_n\mu_n E,$$

а в полупроводнике p -типа —

$$i_E = i_{En} + i_{Ep} = qp_p\mu_p E + qn_p\mu_n E \approx qp_p\mu_p E.$$

Если пренебречь частью проводимости, определяемой неосновными носителями, то для полупроводников n -типа и p -типа можно соответственно написать

$$\sigma_n = qn_n\mu_n \text{ и } \sigma_p = qp_p\mu_p.$$

Следовательно, зная влияние температуры на концентрацию и подвижность носителей заряда, можно оценить температурную зависимость удельной проводимости.

В диапазоне температур, соответствующих области примесной проводимости (область истощения примесей), когда концентрация основных носителей остается практически постоянной, температурные изменения удельной проводимости обусловлены температурной зависимостью подвижности, т. е. электропроводность примесных полупроводников будет уменьшаться с ростом температуры.

В области высоких температур, концентрация электронов и дырок значительно возрастает за счет разрыва ковалентных связей и, несмотря на уменьшение их подвижности, электропроводность примесного полупроводника увеличивается по экспоненциальному закону.

Напомним также, что всегда $\mu_n > \mu_p$ и, следовательно, при одинаковых степенях легирования $\sigma_n > \sigma_p$ и $\sigma_{Ge} > \sigma_{Si}$.

5. ГЕНЕРАЦИЯ И РЕКОМБИНАЦИЯ В ПОЛУПРОВОДНИКАХ

В состоянии термодинамического равновесия процессы генерации и рекомбинации носителей заряда взаимно уравновешены. Но при этом генерация является ведущей в этом единстве и связана с воздействием энергетических факторов: теплового хаотического движения атомов кристаллической решетки (тепловая генерация), поглощенных полупроводником квантов света (световая генерация) и др. Поскольку полупроводник всегда находится под действием всех этих факторов или хотя бы одного ($T \neq 0$), генерация носителей происходит непрерывно. Рекомбинация представляет собой внутреннюю реакцию системы на появление и возрастание числа носителей. Именно рекомбинация, противодействуя накоплению носителей, обуславливает равновесные концентрации электронов n_0 и дырок p_0 .

При воздействии на полупроводник дополнительного внешнего энергетического фактора (свет, сильный электрический ток и др.), из-за генерации новых носителей заряда, их концентрация n и p (неравновесная концентрация) будет превышать равновесную концентрацию на величину Δn (или Δp), которую называют избыточной концентрацией. Таким образом,

$$\Delta n = n - n_0; \quad \Delta p = p - p_0.$$

Избыточная концентрация носителей заряда может возникать в отдельных областях полупроводниковой структуры прибора не только в результате внешних энергетических воздействий, но и за счет различных процессов (инжекция, экстракция, аккумуляция и т. д.), которые могут происходить в полупроводниковых приборах.

В зависимости от источника дополнительной энергии различают:

- 1) тепловую генерацию — происходит при воздействии теплового хаотического движения атомов кристаллической решетки $V_{\text{ген}} = kT$; идет и в собственном и в примесном полупроводни-

- ке. С ростом температуры концентрация возрастает экспоненциально;
- 2) фотогенерацию (световую) — происходит при воздействии поглощенных полупроводником квантов света (это может быть рентгеновское или γ -излучение). Должно выполняться условие $h\nu > \Delta W_3$ в собственном полупроводнике или $h\nu > \Delta W_{\text{пр}} = W_{\text{п}} - W_{\text{д}}$ в примесных полупроводниках;
 - 3) ударную ионизацию — лавинное размножение носителей. В сильных полях $E > E_{\text{кр}}$ ($E_{\text{кр}} \approx 3 \cdot 10^4$ В/см для кремния) на длине свободного пробега электрон проводимости e^- (или дырка проводимости e^+) набирает дополнительную энергию, достаточную для ионизации атомов примеси или собственных атомов полупроводника;
 - 4) полевую генерацию — электрическое поле срывает e^- с внешних оболочек атомов. Возникает при $E > 10^5 \dots 10^6$ В/см.

Если в ограниченный участок объема полупроводника ввести избыточные концентрации электронов и дырок, то в полупроводнике возникнет электрическое поле E , под влиянием которого избыточные заряды будут покидать тот объем, в который они были введены. Изменение их концентрации определяется из уравнения непрерывности, которое для данного случая имеет вид

$$\frac{\partial n}{\partial t} = n\mu_n \frac{\partial E}{\partial x}; \quad \frac{\partial p}{\partial t} = -p\mu_p \frac{\partial E}{\partial x},$$

где $\partial E / \partial x$ — изменение напряженности электрического поля E по геометрической координате x ; μ_n и μ_p — подвижность электронов и дырок.

Решение уравнений позволяет определить разность избыточных концентраций Δp и Δn в любой момент времени:

$$\Delta p - \Delta n = [\Delta p(0) - \Delta n(0)] \cdot e^{-t/\tau_e},$$

где τ_e — время диэлектрической релаксации, $\tau_e = \frac{\epsilon_0 \epsilon}{q(n_0 \mu_n + p_0 \mu_p)}$. Здесь

ϵ_0 — диэлектрическая постоянная воздуха; ϵ — относительная диэлектрическая проницаемость полупроводника.

Итак, переходный процесс имеет аperiодический характер и заканчивается в течение времени $(3 \dots 5) \tau_e$.

Таким образом, если в полупроводник введено разное количество электронов и дырок, то разность концентраций носителей заряда противоположного знака стремится к нулю, уменьшаясь по экспоненциальному закону. Время диэлектрической релаксации не более $\tau_{\varepsilon} \approx 10^{-12}$ с. Поэтому процесс уравнивания зарядов одного знака зарядами другого происходит за очень короткий промежуток времени. Это дает возможность сделать важный теоретический вывод: в однородном полупроводнике, независимо от характера и скорости образования носителей заряда в условиях как равновесной, так и неравновесной концентрации, не могут иметь место существенные объемные заряды в течение времени большего $(3...5) \tau_{\varepsilon}$, за исключением участков малой протяженности. Этот вывод называют условием электронейтральности, или квазиэлектронейтральности полупроводника. Ограничение относительно участков малой протяженности касается участков p - n переходов и поверхностных слоев, которые при рассмотрении полупроводника в целом также можно считать электронейтральным.

Условие электронейтральности для полупроводника, в котором имеются электроны и дырки с концентрациями n и p и ионы акцепторной и донорной примесей с концентрациями $N_{\text{д}}^+$, $N_{\text{а}}^-$, записывают в виде

$$n + N_{\text{а}}^- = p + N_{\text{д}}^+.$$

Различают два механизма обеспечения условия электронейтральности:

- 1) если в полупроводник с электропроводностью определенного типа, например p , ввести некоторое количество дырок (основных носителей), концентрация которых равна $\Delta p(0)$, то они уходят из начального объема, изменяя свою концентрацию в соответствии с выражением

$$\Delta p = \Delta p(0) e^{-t/\tau_{\varepsilon}};$$

- 2) если в полупроводник n -типа ввести дополнительные дырки (неосновные носители), концентрация которых $\Delta p(0)$, то электроны из объема полупроводника под действием электрического поля приходят в область объема, куда были введены дырки, компенсируя заряд неосновных носителей — дырок. В итоге в этом объеме через время $t(3...5)\tau_{\varepsilon}$ окажется дополнительный заряд электронов Δn , равный заряду введенных дырок $\Delta p(0)$, $\Delta n = \Delta p(0) \cdot (1 - e^{-t/\tau_{\varepsilon}})$.

Таким образом, если возмущение было вызвано основными носителями заряда, то рассасывание их произойдет за малый промежуток времени. Если возмущение вызвано неосновными для данного полупроводника носителями заряда, то в течение короткого времени в полупроводнике появится дополнительный заряд основных носителей, компенсирующий заряд неосновных носителей.

Если возмущение, в результате которого появилась дополнительная концентрация носителей заряда в полупроводнике, закончилось, то эти заряды в результате рекомбинации рассасываются, причем их концентрация убывает по экспоненциальному закону

$$\Delta p = \Delta n = \Delta n(t_1) e^{-t/\tau} = \Delta p(t_1) e^{-t/\tau},$$

где $\Delta n(t_1)$ — концентрация носителей заряда в момент прекращения возмущения и окончания процесса нейтрализации, $\Delta n(t_1) = \Delta p(t_1)$; τ — время жизни носителей заряда.

Время жизни носителей заряда $\tau > \tau_e$, поэтому рассасывание заряда происходит значительно дольше, чем его нейтрализация.

Рассмотрим случай линейной рекомбинации, когда в полупроводник с явно выраженной примесной электропроводностью введены неосновные носители заряда в небольшом количестве, тогда появление неравновесных неосновных носителей заряда не вызывает существенного изменения концентрации основных, с которыми происходит рекомбинация. Время жизни при этом оказывается постоянным, а количество носителей заряда, рекомбинировавших в единицу времени в единице объема, пропорционально первой степени избыточной концентрации $-\frac{dn}{dt} = \frac{\Delta n}{\tau_n}$.

Пусть в момент времени $t = 0$ в полупроводнике p -типа создана избыточная концентрация $\Delta n(0)$, которая после окончания действия источника избыточных носителей заряда должна стремиться к нулю при $t \rightarrow \infty$. Решение уравнений при этих условиях имеет вид

$$\Delta n(t) = \Delta n(0) \exp(-t / \tau_n),$$

т. е., при линейной рекомбинации, избыточная концентрация неосновных носителей за время жизни уменьшается в e раз (рис. 29).

Вывод: изменение избыточной концентрации неосновных носителей происходит с постоянной времени близкой к равновесному времени жизни неосновного носителя.

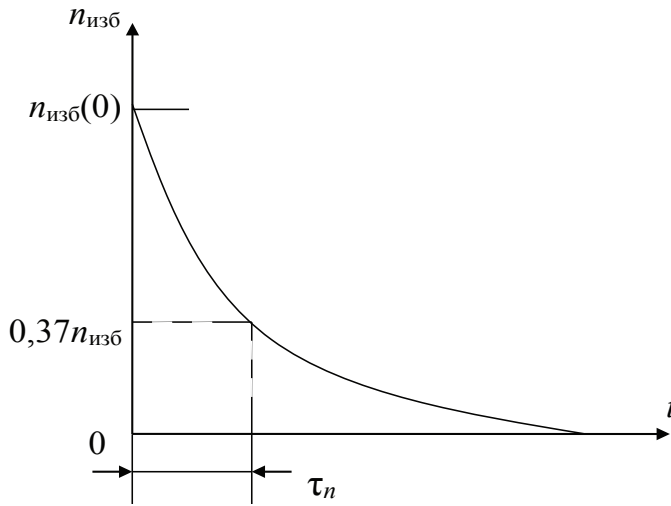


Рис. 29. Изменение избыточной концентрации во времени

Механизмы рекомбинации могут быть различны (рис. 30).

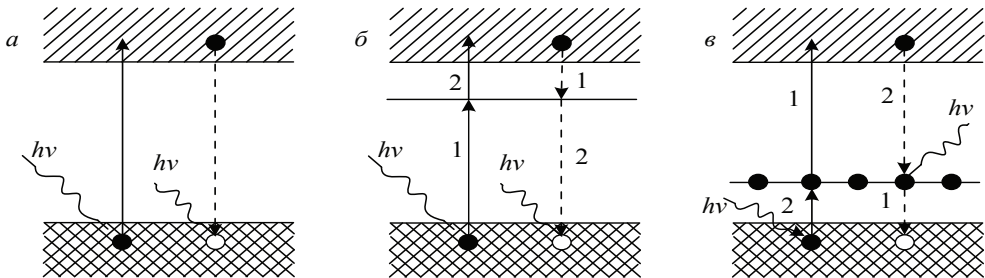


Рис. 30. Механизмы генерации и рекомбинации носителей заряда:

a — межзонная генерация и рекомбинация; *б* — генерация и рекомбинация с участием пустых рекомбинационных ловушек; *в* — генерация и рекомбинация с участием заполненных электронами рекомбинационных ловушек;
 ————— — генерация; — — — — рекомбинация; цифры — этапы процессов генерации и рекомбинации

Межзонная, или непосредственная, рекомбинация происходит при переходе свободного электрона из зоны проводимости в валентную зону на один из свободных энергетических уровней, что соответствует исчезновению пары носителей заряда — свободного электрона и дырки. Однако такой процесс межзонной рекомбинации маловероятен, т. к. свободный электрон и дырка должны оказаться одновременно

в одном и том же месте кристалла. Кроме того, должен выполняться закон сохранения импульса, т. е. рекомбинация электрона и дырки возможна только при одинаковых, но противоположно направленных импульсах электрона и дырки. Поэтому, например, в германии на 10 тыс. рекомбинаций лишь одна происходит в результате межзонной рекомбинации.

В большинстве полупроводников, используемых в настоящее время, рекомбинация происходит через рекомбинационные центры, которые называют рекомбинационными ловушками. Рекомбинационные ловушки — это атомы примесей, различные включения в кристалле, незаполненные узлы кристаллической решетки, трещины и другие дефекты кристаллической структуры (объема или поверхности), энергетические уровни, которые находятся в глубине запрещенной зоны (достаточно далеко как от валентной зоны, так и от зоны проводимости).

Двухэтапный процесс рекомбинации более вероятен, т. к. он не требует одновременного присутствия в данном месте кристалла свободного электрона и дырки. Рекомбинационная ловушка воспринимает количество движения, необходимое для соблюдения закона сохранения импульса, и может забрать часть энергии, освобождаемой в процессе рекомбинации.

Рекомбинация на ловушках (см. рис. 30) носит ступенчатый характер. Для полупроводника p -типа уровень Ферми W_{Fp} вблизи валентной зоны значительно ниже ловушек, почти все ловушки свободны. На первой стадии рекомбинационная ловушка захватывает e^- из зоны проводимости. Это относительно медленный процесс, т. к. он определяется столкновением e^- с ловушками, а концентрация e^- в полупроводнике p -типа мала. На втором этапе e^- переходит с уровня ловушки в валентную зону. Это относительно быстрый процесс, т. к. он определяется столкновением e^+ с ловушками (захватом дырки), а концентрация e^+ в полупроводнике p -типа велика. Скорость рекомбинации и время жизни τ_n определяются 1-й стадией, причем τ_n обратно пропорционально концентрации свободных ловушек. В полупроводнике n -типа, уровень Ферми W_{Fn} лежит значительно выше уровня ловушек и почти все ловушки заняты e^- . Основной стадией рекомбинации, определяющей время жизни, является захват e^+ ловушкой, т. е. переход e^- с уровня ловушки в валентную зону. Второй стадией является переход e^- из зоны проводимости на уровень ловушки. Таким образом, в полупроводнике любого типа проводимости 1-й и основной

стадией рекомбинации, от которой зависит время жизни носителя, является захват неосновного носителя ловушкой. Вторая, более быстрая стадия, — захват основного носителя.

В связи с тем, что на поверхности кристалла перечисленных дефектов значительно больше, чем в объеме, процесс рекомбинации на поверхности должен идти значительно интенсивнее. Его рассматривают и оценивают обычно отдельно, считая поверхностную рекомбинацию разновидностью рекомбинации с участием рекомбинационных ловушек.

В зависимости от того как расходуется энергия, освобождающаяся при рекомбинации электрона и дырки, рекомбинацию можно подразделить на два вида: излучательную и безызлучательную.

Излучательной рекомбинацией (см. рис. 30) называют рекомбинацию, при которой энергия, освобождающаяся при переходе электрона на более низкий энергетический уровень, излучается в виде кванта света (фотона). При безызлучательной (фононной) рекомбинации, избыточная энергия электрона передается кристаллической решетке полупроводника, т. е. избыточная энергия идет на образование фононов — квантов тепловой энергии. Строго говоря, характер излучения зависит от строения зон полупроводника. Если экстремумы у зон совпадают (рис. 31, *а*) (в реальном полупроводнике ширина запрещенной зоны меняется в зависимости от геометрической координаты) и при переходе электрона значение его импульса $p = m_n^* v$ остается постоянным, то энергия ΔW выделяется в виде фотона. При несовпадении экстремумов (рис. 31, *б*) обычно имеет место безызлучательная рекомбинация с выделением фонона.

Время жизни неравновесных носителей заряда зависит от температуры полупроводника. Рассмотрим температурную зависимость времени жизни носителей заряда на примере полупроводника с электропроводностью *n*-типа с рекомбинационными ловушками в верхней половине запрещенной зоны (рис. 32). При очень низкой температуре рекомбинационные ловушки заполнены электронами, т. к. вероятность их заполнения, судя по положению уровня Ферми, намного больше 50 %. При этом 1-й этап рекомбинации (захват дырки рекомбинационной ловушкой) происходит быстро и время жизни оказывается небольшим. С повышением температуры уровень Ферми смещается вниз и находится вблизи энергетических уровней рекомбинационных ловушек. Это означает, что теперь не все ловушки заполнены электронами, т. е.

не все ловушки могут захватить блуждающие по полупроводнику дырки. Поэтому с повышением температуры время жизни растет.

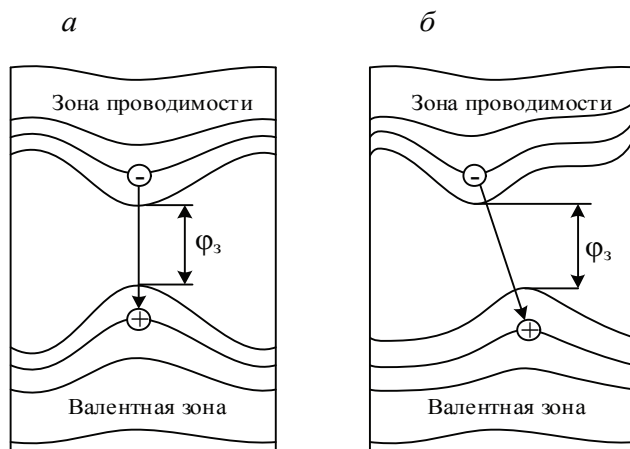


Рис. 31. Процесс межзонной рекомбинации носителей заряда:
а — при совпадении экстремумов; *б* — при несовпадении экстремумов

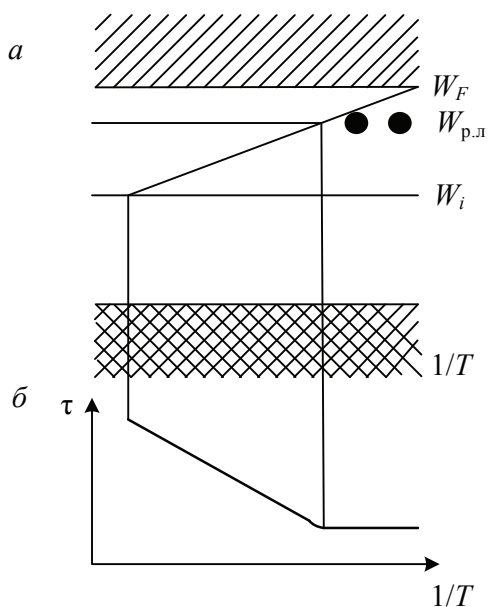


Рис. 32. Температурная зависимости времени жизни:
а — зависимость положения уровня Ферми от температуры
и заполнения уровней рекомбинационных ловушек электронами;
б — зависимость времени жизни от температуры

Кроме того, с ростом температуры увеличивается тепловая скорость неосновных носителей, а значит, уменьшается вероятность их захвата рекомбинационными ловушками.

Необходимо отметить, что рассмотренная температурная зависимость времени жизни справедлива только для полупроводника с рекомбинационными ловушками одного сорта или типа. Если же в полупроводнике будут рекомбинационные ловушки разных типов, создающие в запрещенной зоне несколько различных энергетических уровней, то температурная зависимость времени жизни может быть сложнее представленной на рис. 32.

Время жизни носителей заряда в значительной степени зависит от концентрации в полупроводнике рекомбинационных ловушек, т. к. рекомбинация чаще всего происходит с их помощью. Но кроме рекомбинационных ловушек в запрещенной зоне полупроводника обычно существуют энергетические уровни, которые могут захватывать носители только какого-либо одного типа. Такие уровни называют уровнями ловушек захвата, а дефекты кристаллической решетки, создающие уровни ловушек захвата, — ловушками захвата.

Энергетические уровни ловушек захвата электронов расположены в запрещенной зоне вблизи зоны проводимости, ловушек захвата дырок — вблизи валентной зоны (рис. 33).

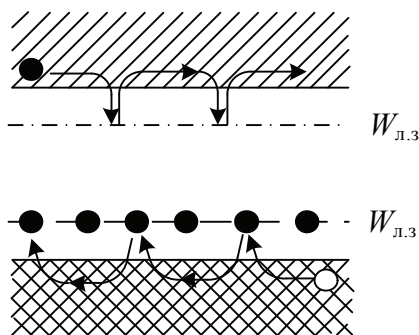


Рис. 33. Расположение энергетических уровней ловушек захвата и схематическое изображение процесса захвата электронов и дырок

Через некоторое время после захвата носителей заряда ловушкой захвата может произойти ионизация этой ловушки, т. е. освобождение носителя заряда. Если время нахождения носителя в ловушке захвата велико или велика концентрация ловушек захвата, то эффектив-

ное время жизни носителя заряда может оказаться значительно больше действительного времени жизни, т. к. находящийся в ловушке захвата носитель не может в это время рекомбинировать. Рекомбинация возможна только после ионизации ловушки захвата или после освобождения носителя заряда.

Таким образом, в запрещенной зоне энергетической диаграммы полупроводника может существовать много различных локальных энергетических уровней, связанных с наличием разных примесей. Часть из них может быть уровнями ловушек захвата, часть — уровнями рекомбинационных ловушек. При различной степени отклонения от термодинамического равновесия, роль, выполняемая отдельными ловушками, может изменяться, т. е. ловушки захвата могут стать рекомбинационными ловушками, и наоборот. Чтобы установить количественный критерий отличия этих энергетических уровней, введены понятия демаркационных уровней, для которых вероятность ионизации с образованием носителя одного знака равна вероятности захвата носителя заряда противоположного знака.

В неравновесном состоянии распределение свободных электронов и дырок уже не соответствует распределению этих носителей заряда при термодинамическом равновесии. Поэтому в неравновесном состоянии распределение свободных электронов по энергетическим уровням характеризуется своим квазиуровнем Ферми для электронов, распределение дырок — своим квазиуровнем Ферми для дырок, которые имеют тот же смысл для полупроводника в неравновесном состоянии, что и уровень Ферми в условиях термодинамического равновесия. Чем больше неравновесные концентрации свободных электронов и дырок отклоняются от своих равновесных значений, тем больше отличается положение квазиуровней Ферми для электронов и дырок от положения уровня Ферми в условиях термодинамического равновесия.

При качественном рассмотрении различных зависимостей параметров полупроводниковых приборов можно в первом приближении считать демаркационные уровни совпадающими с соответствующими квазиуровнями Ферми (рис. 34). Уровни, лежащие выше электронного демаркационного уровня или выше квазиуровня Ферми для электронов, являются уровнями захвата электронов. Вероятность их заполнения электронами меньше 50 %, что соответствует большей вероятности (более 50 %) переброса электронов с этих уровней в зону проводимости в результате тепловой генерации.

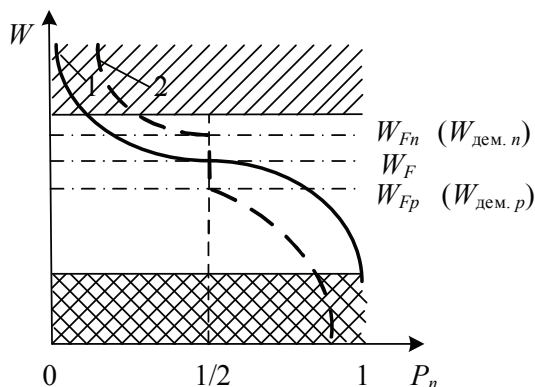


Рис. 34. Вероятность заполнения энергетических уровней в состоянии термодинамического равновесия (кривая 1) и в неравновесном состоянии (кривая 2) и расположение квазиуровней Ферми или демаркационных уровней для электронов и дырок

Аналогично этому уровни лежащие ниже дырочного демаркационного уровня или квазиуровня Ферми для дырок, являются уровнями захвата дырок. Для уровней, расположенных между электронными и дырочными демаркационными уровнями или между квазиуровнями Ферми, характерна большая вероятность заполнения, с одной стороны, электронами, а с другой стороны, дырками. В действительности сумма вероятностей заполнения какого-либо уровня электроном и дыркой должна быть равна единице. Поэтому следует считать, что заполнение электронами и дырками всех уровней, расположенных между демаркационными уровнями $W_{дем\ n}$ и $W_{дем\ p}$, одинаково. В связи с этим уровни, расположенные между электронным и дырочным демаркационными уровнями, следует считать уровнями рекомбинационных ловушек.

6. ДИФфуЗИОННЫЙ ТОК. ЗАКОНЫ ДВИЖЕНИЯ НОСИТЕЛЕЙ ЗАРЯДА В ПОЛУПРОВОДНИКАХ

Диффузия (лат. *diffusio*) — распространение, растекание, рассеивание, — это перенос частиц разной природы, обусловленный хаотическим тепловым движением молекул (атомов) при наличии градиента концентрации частиц. Диффузия имеет место как в газах и жидкостях, так и в твердых телах и не связана с электрическим зарядом частиц. Диффузия происходит в направлении падения концентрации вещества и ведет к выравниванию концентрации вещества в первоначально неоднородной среде. Эту аналогию можно распространить и на явления, происходящие в результате неравномерного распределения концентрации носителей заряда в объеме полупроводника в отсутствие градиента температуры.

В полупроводниках может возникать диффузионный ток, которым называется направленное перемещение носителей заряда в сторону меньшей концентрации за счет собственной энергии (т. е. за счет теплового движения). Поясним появление диффузионного тока в одномерном случае.

Если носители заряда распределены равномерно по полупроводнику, то их концентрация является равновесной. Под влиянием каких-либо внешних воздействий в разных частях полупроводника, концентрация может быть неодинаковой. Например, если часть полупроводника подвергнуть действию излучения, то в ней усилится генерация пар носителей и возникнет дополнительная концентрация носителей, называемая избыточной. Следовательно, избыточные носители заряда не находятся в тепловом равновесии с решеткой, поэтому они называются неравновесными. В отличие от равновесных, они могут неравномерно распределяться по объему полупроводника.

Неравномерное распределение неравновесных носителей зарядов сопровождается их диффузией в сторону меньшей концентрации. Это движение носителей зарядов (электронов и дырок) обуславливает прохождение электрического тока, называемого диффузионным $I_{\text{диф}}$. Такой ток, как и ток проводимости, может быть электронным и дырочным.

Главной характеристикой диффузии служит плотность диффузионного тока $i_{\text{диф}}$ — количество носителей заряда, переносимых в единицу времени через единицу площади поверхности, перпендикулярной направлению переноса. Плотности диффузионных токов определяются по закону Фика

$$i_{\text{диф } n} = qD_n \frac{dn}{dx},$$

$$i_{\text{диф } p} = -qD_p \frac{dp}{dx},$$

где D_n и D_p — коэффициенты диффузии; dn/dx и dp/dx — градиенты концентрации.

Градиент концентрации характеризует, насколько резко меняется концентрация вдоль оси x (одномерный случай), т. е. каково изменение концентрации n или p на единицу длины. Если разности концентраций нет, то $\Delta n = 0$ или $\Delta p = 0$ и ток диффузии не возникает. Чем больше изменение концентрации Δn или Δp на данном расстоянии Δx , тем больше ток диффузии.

Знак «минус» в формуле плотности дырочного диффузионного тока указывает на то, что дырочный ток направлен в сторону уменьшения концентрации дырок. Это поясняет рис. 35, на котором показано, что если концентрация дырок p возрастает с увеличением координаты x , то дырки двигаются в сторону, противоположную положительному направлению оси x . Следовательно, дырочный ток в этом случае надо считать отрицательным (рис. 36).

Коэффициенты диффузии D_n , D_p — количество носителей заряда, пересекающих в единицу времени единичную площадку, перпендикулярную выбранному направлению, при градиенте концентрации в этом направлении, равном единице. Это параметры диффузионного движения.

Параметры дрейфового и диффузионного движения связаны между собой соотношением Эйнштейна

$$D_n = (kT/q)\mu_n = \varphi_T \mu_n;$$

$$D_p = (kT/q)\mu_p = \varphi_T \mu_p,$$

где D_n , μ_n — коэффициент диффузии и подвижность электронов;
 D_p , μ_p — коэффициент диффузии и подвижность дырок.

Эти соотношения строго выполняются лишь для невырожденных полупроводников в условиях равновесия.

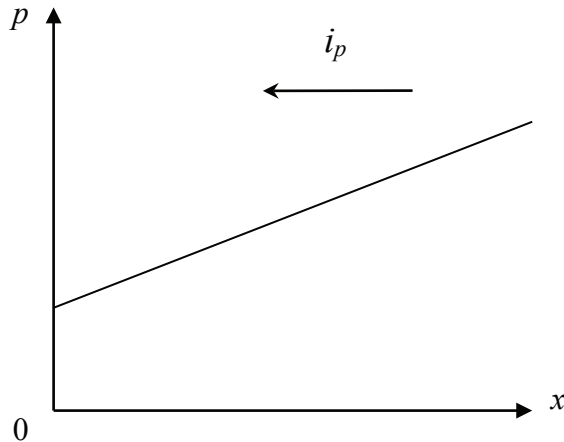


Рис. 35. Движение дырок при наличии разности концентраций



Рис. 36. Диффузионный ток в полупроводниках n - и p -типа

Соотношение Эйнштейна имеет простой физический смысл: независимо от причины, вызвавшей направленное движение, свободные носители встречают на своем пути одни и те же неоднородности (препятствия), при взаимодействии с которыми происходит рассеяние. Поэтому между основными параметрами дрейфового и диффузионного движений μ и D существует некая пропорциональность. Коэффициент пропорциональности $\varphi_T = kT/q$, В, называется тепловым по-

тенциалом. Поэтому все описанные выше зависимости подвижности носителей заряда могут быть использованы и для коэффициентов диффузии с учетом множителя φ_T (рис. 37). Коэффициент диффузии в основном зависит от температуры, типа носителя (табл. 3), материала полупроводника, степени легирования, центров рассеяния.

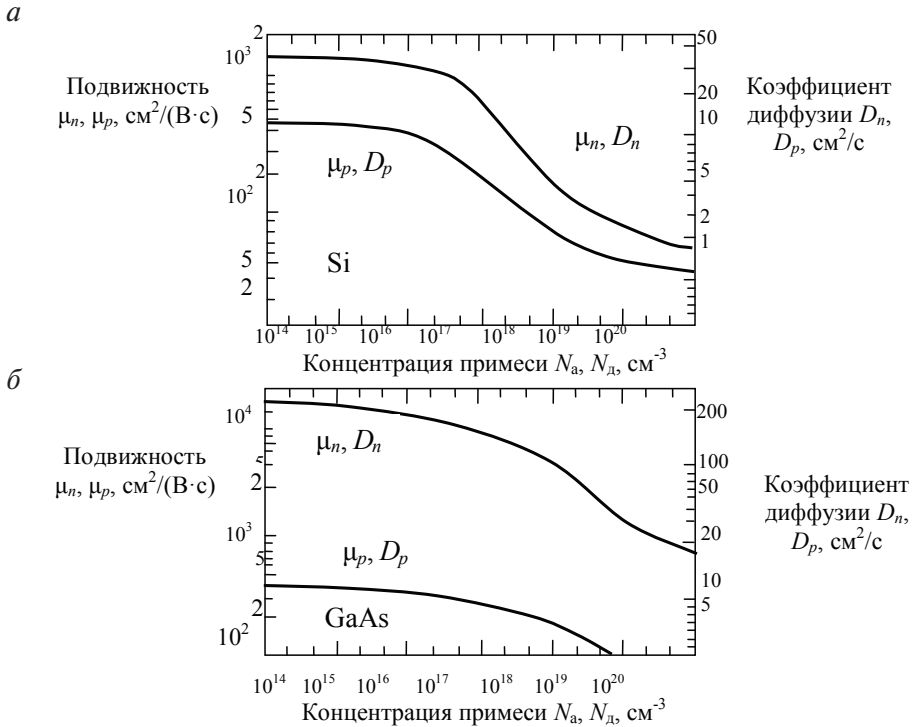


Рис. 37. Зависимости подвижности (*а*) и коэффициента диффузии (*б*) от концентрации примесей для полупроводников Si и GaAs при комнатной температуре

Таблица 3

Коэффициент диффузии носителей в чистых полупроводниках при комнатной температуре

Коэффициент диффузии $\text{см}^2/\text{с}$	Ge	Si	GaAs	InSb
D_n	100	36	260	до 1750
D_p	45	13	12	17

Поскольку подвижность для электронов и дырок различна и зависит от материала полупроводника, как было рассмотрено выше, отличаются по величине и коэффициенты диффузии $D_n > D_p$ и $D_{Ge} > D_{Si}$.

При низких температурах, когда преобладает рассеяние на примесях, коэффициент диффузии $D \sim T^{5/2}$, а при высоких температурах $D \sim T^{-1/2}$. Различие между температурной зависимостью подвижности и коэффициентом диффузии приводит к тому, что, с ростом температуры, относительная роль диффузионного движения увеличивается.

Зная характер изменения подвижностей и соотношения Эйнштейна, можно подучить зависимость коэффициентов диффузии от температуры

$$D = D_0 \left(\frac{T_0}{T} \right)^{1/2},$$

где D_0 — коэффициент диффузии при температуре T_0 .

Аналогично этому можно определить зависимость коэффициентов диффузии от концентрации примесей

$$D = D_0 \left(\frac{N_0}{N} \right)^{1/3},$$

где D_0 — коэффициент диффузии при концентрации примесей N_0 .

Если за счет какого-то внешнего воздействия в некоторой части полупроводника создана избыточная концентрация носителей, а затем внешнее воздействие прекратилось, то избыточные носители будут рекомбинировать и распространяться путем диффузии в другие части полупроводника.

Избыточная концентрация начнет убывать по экспоненциальному закону, которая показана на рис. 38, а для электронной концентрации.

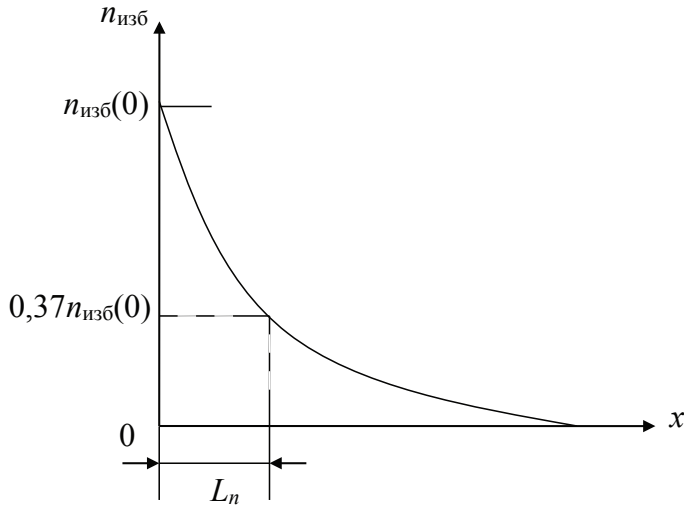
Расстояние L_n , на котором избыточная концентрация неравновесных носителей уменьшается в e (2,72) раз, т.е. становится равной 0,37 от первоначального значения n_0 , называют диффузионной длиной в полупроводнике p -типа. Она характеризует убывание избыточной концентрации в пространстве.

Скорость рекомбинации неравновесных носителей (рис. 38, б) пропорциональна избыточной концентрации дырок ($p_n - p_{n0}$) и электронов ($n_p - n_{p0}$)

$$\frac{dp}{dt} = -\frac{p_n - p_{n0}}{\tau_p}, \quad \frac{dn}{dt} = -\frac{n_p - n_{p0}}{\tau_n},$$

где τ_p — время жизни дырок; τ_n — время жизни электронов.

a



б

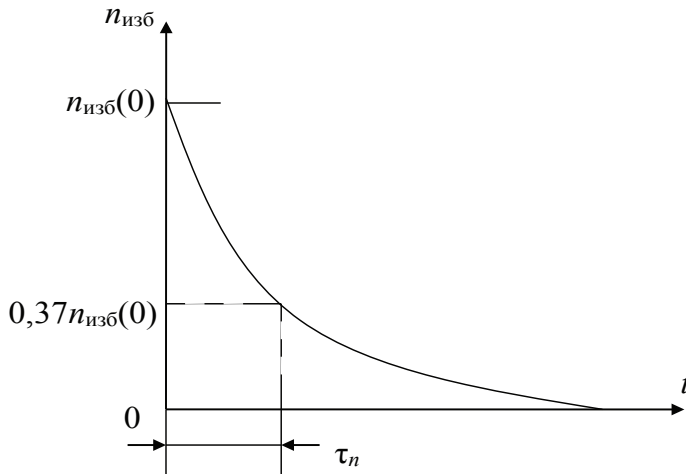


Рис. 38. Изменение избыточной концентрации в пространстве (*a*) и во времени (*б*)

Время, в течение которого избыточная концентрация уменьшится в e (2,72) раз, т.е. станет равна 0,37 от первоначального значения $n_{\text{изб}}(0)$, называют временем жизни неравновесных носителей τ_n . По этой величине характеризуют процесс убывания избыточной концентрации во времени. Рекомбинация неравновесных носителей происходит внутри полупроводника и на его поверхности и сильно зависит от примесей, а также от состояния поверхности. Значения τ_n для германия и кремния в различных случаях могут быть долей микросекунды до сотен микросекунд и более.

Таким образом, убывание избыточной концентрации происходит во времени и пространстве, поэтому величины τ_n и L_n оказываются связанными друг с другом по следующей зависимости:

$$L_n = \sqrt{D_n \tau_n}.$$

На основании уравнения можно сделать вывод, что с увеличением температуры уменьшается D_n , резко возрастает τ_n и возрастает диффузионная длина.

Если в полупроводнике существует и электрическое поле, и градиент концентрации носителей, то полный проходящий ток будет иметь дрейфовую и диффузионную составляющие. В таком случае плотности токов рассчитываются по следующим уравнениям:

$$i_n = qn\mu_n E + qD_n \frac{dn(x)}{dx};$$

$$i_p = qp\mu_p E - qD_p \frac{dp(x)}{dx}.$$

Таким образом, плотность общего тока i в любой точке неоднородного полупроводника в любой момент времени будет определяться по уравнению

$$i = i_n + i_p = q(n\mu_n + p\mu_p)E + q(D_n \frac{dn(x)}{dx} - D_p \frac{dp(x)}{dx}).$$

Следовательно, для определения тока необходимо знать концентрацию носителей и напряженность поля. В общем случае концентрации p и n зависят от двух переменных: координаты x и времени t . Поэтому для определения токов нужно предварительно найти функции $p(x, t)$ и $n(x, t)$. Эти функции являются решениями так называемых уравнений непрерывности потока, которым в любой момент времени подчиняется движение носителей.

При отсутствии внешних факторов (свет, радиация и т. п.) уравнения непрерывности для дырок и электронов можно записать в виде

$$\frac{dp}{dt} = -\frac{p-p_0}{\tau_p} + D_p \frac{\partial^2 p}{\partial x^2} - \mu_p E \frac{\partial p}{\partial x} - p \mu_p \frac{\partial E}{\partial x},$$

$$\frac{dn}{dt} = -\frac{n-n_0}{\tau_n} + D_n \frac{\partial^2 n}{\partial x^2} - \mu_n E \frac{\partial n}{\partial x} - n \mu_n \frac{\partial E}{\partial x}.$$

Если напряженность E меняется вдоль оси x (если в полупроводнике имеется существенный объемный заряд), приходится дополнительно привлекать уравнение Пуассона, которое в одномерном случае имеет вид

$$\frac{dE}{dx} = \frac{\lambda}{\epsilon_0 \epsilon},$$

где λ — плотность заряда; ϵ_0 — электрическая постоянная; ϵ — относительная диэлектрическая проницаемость.

На основании анализа последних уравнений можно сделать вывод, что изменение концентраций носителей заряда в полупроводнике с течением времени происходит из-за рекомбинации (первые слагаемые правых частей), перемещения вследствие диффузии (вторые слагаемые) и дрейфа (третьи и четвертые слагаемые).

Решение системы уравнений в общем виде невозможно. В каждом конкретном случае приходится вводить те или иные упрощения.

Задачи к ПРАКТИЧЕСКОМУ ЗАНЯТИЮ № 2

- 1) В собственном германии концентрация атомов равна $4,5 \cdot 10^{28} \text{ м}^{-3}$. При $T = 300 \text{ К}$ один из каждых $2 \cdot 10^9$ атомов ионизирован. Определить: а) удельную проводимость собственного германия; б) удельную проводимость германия при $T = 300 \text{ К}$, легированного элементом V группы, если на каждые 10^8 атомов германия приходится один атом примеси.
- 2) К стержню из арсенида галлия длиной 50 мм приложено напряжение 50 В. За какое время электрон пройдет через весь образец, если подвижность электронов $\mu_n = 0,9 \text{ м}^2/(\text{В} \cdot \text{с})$?

- 3) В полупроводниковом кристалле под действием света образуется равномерно распределенная избыточная концентрация носителей заряда Δn . Равновесная концентрация неосновных носителей заряда составляет $2,5 \cdot 10^{20} \text{ м}^{-3}$, а начальная скорость уменьшения концентрации равна $2,8 \cdot 10^{24} \text{ с}^{-1}$. Определить: а) время жизни неосновных носителей заряда; б) значение Δn через 2 мс после выключения источника света.
- 4) На сколько изменится коэффициент диффузии электронов в невырожденном полупроводнике при повышении температуры на 10 %, если подвижность электронов изменяется пропорционально $T^{-3/2}$?
- 5) По истечении времени $t_1 = 10^{-4} \text{ с}$ после прекращения генерации электронно-дырочных пар, равномерной по объему полупроводника, избыточная концентрации носителей заряда оказалась в 10 раз больше, чем в момент $t_2 = 10^{-3} \text{ с}$. Определить время жизни неравновесных носителей заряда, считая его постоянным, независимым от интенсивности возбуждения.

Задания: решение задач 7–9 раздела «Однородный полупроводник» (приложение «Домашние задания»).

7. ПРИМЕНЕНИЕ ОДНОРОДНЫХ ПОЛУПРОВОДНИКОВ

Свойство полупроводников изменять свою электропроводность под действием внешних возбуждений используется при построении ряда простейших полупроводниковых приборов (беспереходных), так называемых полупроводниковых резисторов [7].

Полупроводниковые резисторы — это множество видов резисторов, изготовленных из различных полупроводниковых материалов и использующих зависимость их электрического сопротивления от разнообразных воздействующих на резистор факторов. Выделяют следующие типы полупроводниковых резисторов:

- 1) варисторы (зависимость R от напряжения U);
- 2) терморезисторы (зависимость R от температуры T);
- 3) фоторезисторы (зависимость R от светового потока Φ);
- 4) магниторезисторы (зависимость R от магнитного поля B);
- 5) тензорезисторы (зависимость R от механического давления P).

Условные обозначения полупроводниковых резисторов приведены на рис. 39.

Наличие полупроводниковых резисторов с таким широким спектром зависимостей позволяет использовать их в разрабатываемой радиоэлектронной аппаратуре для решения множества разнообразных задач:

- 1) в качестве датчиков для измерения соответствующего параметра (U , T , Φ , B , P);
- 2) в устройствах стабилизации параметров объектов;
- 3) в системах сигнализации и защиты от перегрузок;
- 4) в системах регулирования физических величин;
- 5) в системах преобразования сигналов.

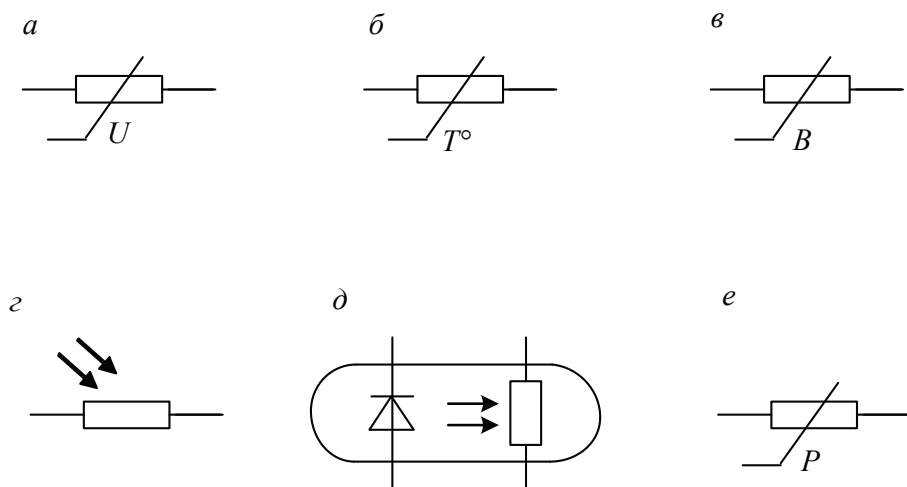


Рис. 39. Условные графические обозначения полупроводниковых резисторов:

а — варистор; *б* — терморезистор; *в* — магниторезистор;
г — фоторезистор; *д* — оптоэлектронная пара со светодиодом и фоторезистором; *е* — тензорезистор

7.1. ВАРИСТОРЫ

Варистор — это полупроводниковый резистор, сопротивление которого нелинейно зависит от приложенного как положительного, так и отрицательного напряжения [7]. Варистор имеет два вывода.

Варисторы достаточно широко используются, прежде всего как элементы защиты от перенапряжений, более того, в силу симметричности высоконелинейной вольт-амперной характеристики, при уникально высокой импульсной устойчивости, оксидно-полупроводниковые варисторы в настоящее время являются практически единственным реальным и массовым быстродействующим средством защиты сложных и дорогостоящих полупроводниковых систем различного назначения.

Вольт-амперная характеристика (ВАХ) варистора нелинейная и симметричная (рис. 40).

Основными материалами, используемыми для изготовления варисторов, являются карбид кремния и оксид цинка.

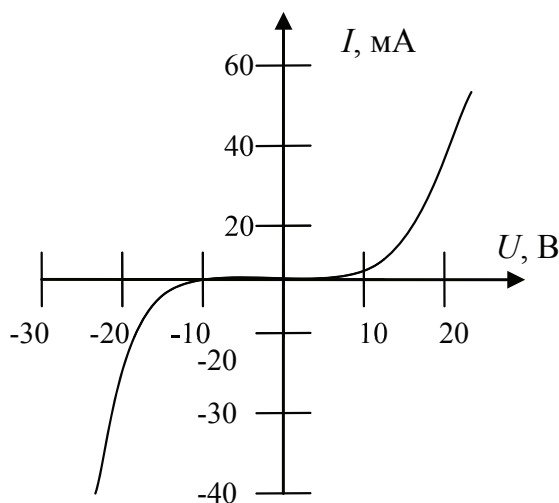


Рис. 40. Вольт-амперная характеристика варистора

Принцип действия варисторов на основе карбида кремния

Чтобы получить такую вольтамперную характеристику $I(U)$, варисторы изготавливают в основном из карбида кремния SiC , порошкообразные зерна которого размером 20–180 мкм перемешивают с 10–40 % связующего диэлектрического материала — глины, керамики, прессуют и обжигают при высокой температуре. В результате этого варистор изнутри представляет собой конгломерат зерен с самой разной величиной зазоров и площадей соприкосновения.

Нелинейность ВАХ данного типа варистора обусловлена увеличением в сильных электрических полях проводимости поверхностных потенциальных барьеров или окисных пленок на кристаллах, а также увеличением проводимости точечных контактов между кристаллами из-за разогрева в связи с выделяющейся на них мощностью.

Поскольку толщина поверхностных потенциальных барьеров и окисных пленок на кристаллах карбида кремния мала, там могут возникать сильные электрические поля даже при малых напряжениях на варисторе, что приводит к туннелированию носителей заряда сквозь потенциальные барьеры или сквозь тонкие оксидные пленки. Таким образом, при малых напряжениях на варисторе, нелинейность ВАХ связана с зависимостью проводимости поверхностных потенциальных барьеров и окисных пленок от величины напряжения.

При больших напряжениях на варисторе и больших токах, проходящих через варистор, плотность тока в точечных контактах оказывается очень большой. Все напряжение, приложенное к варистору, падает на точечных контактах. Поэтому удельная мощность, выделяющаяся в точечных контактах, достигает таких значений, которые нельзя не учитывать. Разогрев точечных контактов приводит к уменьшению их сопротивления и нелинейности ВАХ.

При мелкозернистой структуре эти механизмы практически не зависят от полярности приложенного напряжения, ВАХ варистора получается симметричной.

Сопротивление точечных контактов определяется сопротивлением растекания, т. е. сопротивлением малых активных областей полупроводника под точечными контактами. Из-за малости активных областей, их разогрев практически не приводит к повышению температуры всего варистора. Кроме того, малые объемы активных областей обеспечивают малую инерционность тепловых процессов. Теоретические расчеты показывают, что тепловая постоянная времени активных областей может составлять 10,6–1,7 с. Считая разогрев активных областей одним из основных процессов, приводящих к нелинейности ВАХ в рабочем диапазоне напряжений и токов для варистора, можно получить ряд важных зависимостей и характеристик варистора.

Температурная зависимость удельной проводимости полупроводников соответствует уравнению $\gamma = \gamma_{\infty} e^{\frac{B}{T}}$. Сопротивление растекания двух контактирующих кристаллов будет

$$R_{\text{кон}} \approx \frac{2}{\pi d \gamma} = \frac{2}{\pi d \gamma_{\infty}} e^{\frac{B}{T}},$$

где d — диаметр точечного контакта; B — коэффициент температурной чувствительности поверхностных слоев кристалла.

В таком случае статическое сопротивление варистора, состоящего из N параллельно включенных цепочек, имеющих в свою очередь N' последовательно включенных контактирующих кристаллов, рассчитывается как $R = \frac{N'}{N} R_{\text{кон}} = \frac{2N'}{N\pi d \gamma_{\infty}} e^{\frac{B}{T}}$.

Уравнение теплового баланса для активных областей варистора будет

$$U^2 / R = I^2 R = H(T - T_0),$$

где H — коэффициент рассеяния активных областей; T — температура активных областей; T_0 — температура среды, окружающей активные области.

ВАХ приближенно может быть представлена как

$$U = CI^\alpha, \quad I = BU^\beta,$$

где U, I — напряжение и ток варистора; C, B, α, β — коэффициенты, причем $\alpha = 1/\beta$, $C = B^{-1/\beta} = B^{-\alpha}$.

Коэффициент нелинейности β характеризует нелинейность ВАХ варистора. Для линейных резисторов он равен единице, для нелинейных — существенно больше единицы, причем с увеличением β нелинейность ВАХ возрастает. Выражение для вычисления коэффициента нелинейности можно получить из соотношения между статическим и динамическими сопротивлениями варистора в определенной точке ВАХ

$$\beta = \frac{R_{\text{ст}}}{r} = \frac{U}{I} \frac{dI}{dU}.$$

ВАХ варистора в двойном логарифмическом масштабе (рис. 41) будет

$$\lg U = \frac{1}{\beta} \lg I + \lg C.$$

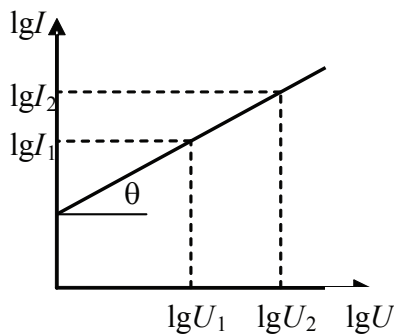


Рис. 41. ВАХ варистора в двойном логарифмическом масштабе

Коэффициент нелинейности в этом случае будет численно равен котангенсу угла наклона характеристики к оси токов $\beta = \text{ctg } \theta = (\lg I_2 - \lg I_1) / (\lg U_2 - \lg U_1)$.

Для вычисления зависимости сопротивления варистора от протекающего по нему тока или приложенного напряжения пользуются соотношением

$$R = CI^{\alpha-1}/B.$$

Иногда в практических расчетах неудобно пользоваться аналитическими выражениями ВАХ. В этом случае в достаточно широком диапазоне напряжений можно пользоваться эмпирическим уравнением $I = \sigma_0 U e^{2a\sqrt{U}}$, где σ_0 — электрическая проводимость рабочего тела варистора в слабых полях; a — постоянная.

В этом случае коэффициент нелинейности β и постоянная B примут следующий вид:

$$\beta = 1 + b\sqrt{U} / 2, \quad B = \sigma_0 U^{-b\sqrt{U}/2} e^{b\sqrt{U}}.$$

Расчетные зависимости β от внешнего напряжения представлены на рис. 42.

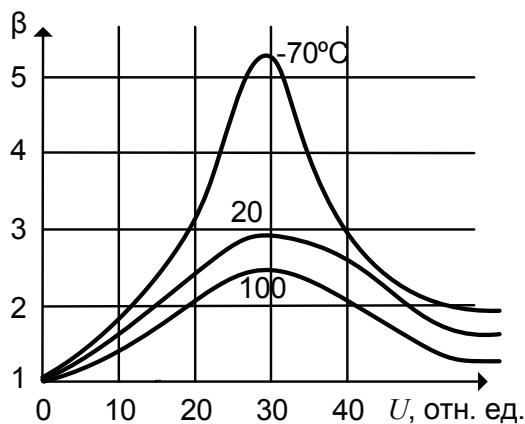


Рис. 42. Расчетная зависимость коэффициента нелинейности от напряжения при различных температурах ($B = 600$ К)

Основные параметры варисторов

Перечислим основные параметры варисторов.

1. Классификационный ток $I_{кл}$ — ток, при котором определяются основные параметры варистора, соответствует рабочей области ВАХ. У большинства варисторов $I_{кл}$ лежит в пределах 2–20 мА, у высоковольтных — равен 0,05 мА.

2. Классификационное напряжение $U_{кл}$ — это напряжение на варисторе, соответствующее классификационному току $I_{кл}$. Разработано множество варисторов с $U_{кл}$ в диапазоне от 5 В до 25 кВ.
3. Асимметрия ВАХ $\delta_U, \%$, оценивается на уровне классификационного тока и обычно не превышает $\pm (5...20) \%$

$$\delta_U = \frac{U_{пр} - U_{обр}}{U_{пр}} \cdot 100\% \text{ при заданном } I_{кл}.$$

4. Статическое сопротивление R — значение сопротивления варистора в заданном режиме при постоянных величинах тока и напряжения $R = U/I$.
5. Динамическое (дифференциальное) сопротивление r — сопротивление варистора малому переменному току

$$r = \frac{dU}{dI} = \frac{\Delta U}{\Delta I}.$$

Динамическое сопротивление r может быть определено по наклону ВАХ варистора в рассматриваемой рабочей точке.

6. Коэффициент нелинейности β — это отношение статического сопротивления к динамическому в рабочем режиме

$$\beta = \frac{R}{r} = \frac{U}{I} \frac{dI}{dU} = \frac{T^2 - BT + BT_0}{T^2 + BT - BT_0},$$

где T — температура активных областей; T_0 — температура среды, окружающей активные области.

Максимальное значение коэффициента нелинейности определяется по формуле

$$\beta_{\max} = \frac{4T_0 + B}{4T_0 - B}.$$

Из соотношения следует, что при $B < 4 T_0$ должны наблюдаться нелинейные ВАХ с положительным коэффициентом нелинейности. Максимальный коэффициент нелинейности должен быть равен бесконечности при $B = 4 T_0$. Если $B > 4 T_0$, то у варистора должно наблюдаться отрицательное дифференциальное сопротивление, коэффициент нелинейности тоже будет отрицательным.

На основе этих расчетов можно заключить, что температура активных областей варистора может превышать температуру

окружающей среды на несколько сотен градусов. Следовательно, для изготовления варисторов со стабильными параметрами необходим термостойкий материал. Именно поэтому используют карбид кремния, кроме того, он является очень дешевым.

Основными примесями в техническом карбиде кремния являются азот и алюминий. Энергия ионизации этих примесей в карбиде кремния невелика (особенно при большой концентрации основных и компенсирующих примесей), соответственно мала и величина коэффициента температурной чувствительности B .

Увеличение температуры окружающей среды должно привести к уменьшению коэффициента нелинейности и незначительному сдвигу максимума кривой $\beta = f(U)$ по напряжению (см. рис. 42).

В справочнике [10] величины β приводятся для $U_{кл}$. В среднем в рабочей области напряжений у варисторов из карбида кремния $\beta = 3 \dots 6$, но имеются варисторы с $\beta \geq 2$ (низковольтные) и с $\beta = 5 \dots 10$ (высоковольтные).

7. Диапазон рабочих температур варисторов лежит в пределах от $T_{\min} = -60 \dots -40^\circ \text{C}$ до $T_{\max} = +70 \dots +125^\circ \text{C}$.

В связи с нелинейностью ВАХ следует различать температурный коэффициент статического сопротивления, измеренный при постоянном напряжении или токе, а также температурные коэффициенты напряжения и тока варистора.

8. Температурные коэффициенты статического сопротивления составят

$$\text{TKR}_{|U=\text{const}} = \frac{1}{R} \frac{\partial R}{\partial T} \Big|_{U=\text{const}} = \beta \text{TKR}_{|I=\text{const}},$$

при малых напряжениях на варисторе, когда $\beta = 1$, $\text{TKR}_{|U \approx 0} = -\frac{B}{T^2}$.

$$\text{TKR}_{|I=\text{const}} = \frac{1}{R} \frac{\partial R}{\partial T} \Big|_{I=\text{const}}.$$

9. Температурные коэффициенты напряжения и тока будут

$$\text{TKU}_{|I=\text{const}} = \frac{1}{U} \frac{\partial U}{\partial T} \Big|_{I=\text{const}},$$

$$\text{TKI}_{|U=\text{const}} = \frac{1}{I} \frac{\partial I}{\partial T} \Big|_{U=\text{const}} = -\beta \text{TKU}_{|I=\text{const}}.$$

Из приведенных уравнений следует, что у варисторов с большим значением максимального коэффициента нелинейности должны быть большие по абсолютной величине значения температурных коэффициентов сопротивления, тока и напряжения. Между максимальным коэффициентом нелинейности и температурными коэффициентами должна быть взаимосвязь, т. к. они зависят от коэффициента температурной чувствительности B ,

$$TKI|_{U=\text{const}} = \frac{4\beta_{\max}(\beta_{\max} - 1)}{(\beta_{\max} + 1)T_0}.$$

Такая взаимосвязь наблюдается при экспериментах, но расчет температурных коэффициентов по экспериментальным данным дает несколько завышенные результаты. Это можно объяснить наличием добавочных сопротивлений, включенных последовательно и параллельно активным областям варистора, повышенной температурой активных областей варистора по сравнению с температурой всего варистора.

У варисторов, выпускаемых отечественной промышленностью, в диапазоне температур $-40...+100\text{ }^{\circ}\text{C}$

$$-TKR|_{U=\text{const}} = TKI|_{U=\text{const}} < 7 \cdot 10^{-3} \text{ K}^{-1},$$

$$-TKR|_{I=\text{const}} = -TKU|_{I=\text{const}} < 1,4 \cdot 10^{-3} \text{ K}^{-1}.$$

10. Допустимая рассеиваемая мощность $P_{\text{РАСС}}$ — мощность, при которой варистор сохраняет свои параметры в заданных технических условиях пределах в течение срока службы. Существуют варисторы с $P_{\text{РАСС}}$ от 10 мВт до 3 Вт.
11. Наибольшая начальная емкость C_0 имеет значение от единиц до сотен пикофарад у различных типов варисторов (рис. 43). Значение определяется только в области токов утечки, поскольку только там варистор проявляет диэлектрические свойства. Емкость зависит от габаритов варистора и материала.
12. Время запаздывания (быстродействие) ограничения перенапряжения при крутых фронтах импульсов напряжения колеблется в пределах 0,5...50 нс.

Эффект запаздывания объясняется индуктивностью выводов. Время срабатывания варисторов можно определить по графику, представленному на рис. 44.

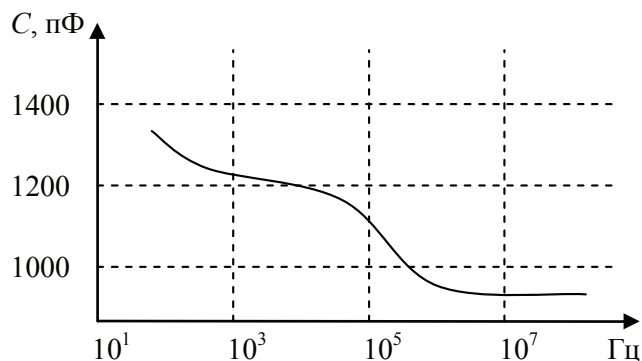


Рис. 43. Частотная зависимость емкости варистора

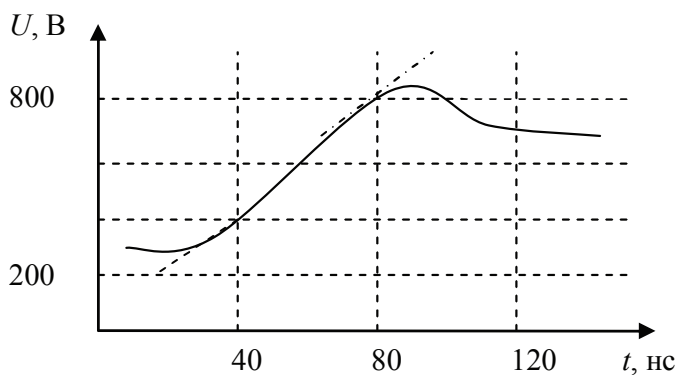


Рис. 44. Зависимость напряжения срабатывания от времени срабатывания варисторов

Варисторы на основе оксида цинка

Представление о механизме проводимости оксидно-цинковых варисторов менялось с углублением знаний об электрофизических свойствах оксидно-цинковой керамики.

Исходным материалом для таких варисторов является оксид цинка ZnO , легированный примесями висмута, кобальта и других элементов. Варисторы из оксидных полупроводников делают методом керамической технологии, которая имеет ряд особенностей, отличающих ее от технологии производства электроизоляционных изделий. Особенности эти обусловлены тем, что нелинейность ВАХ варисторов из оксидных полупроводников связана не со свойствами основной составляющей полупроводниковой керамики — кристаллитами,

а со свойствами межкристаллитных прослоек и потенциальных барьеров на поверхности кристаллитов. Поэтому, кроме традиционных требований обеспечения достаточной плотности с минимальной пористостью обожженного материала, при изготовлении варисторов надо обеспечить высокоомность межкристаллитного слоя, сопротивление которого должно превышать сопротивление объема кристаллитов.

Несмотря на то что первые исследования нелинейных ВАХ кристаллов оксида цинка были проведены О. В. Лосевым еще в 1920-х гг., а технология первых варисторов из оксидных полупроводников была разработана в 1960-х гг., до сих пор нет четкого понимания принципа действия варисторов. Поэтому технология их изготовления имеет эмпирический характер.

Отличием варисторов из оксидных полупроводников от варисторов из карбида кремния является возможность относительно просто получить большую нелинейность ВАХ. Коэффициент нелинейности варисторов из оксида цинка может составлять несколько десятков. Отрицательной особенностью варисторов из оксидных полупроводников является их меньшая стабильность как при работе, так и при хранении, именно поэтому варисторы из оксидных полупроводников до сих пор не получили широкого распространения.

В свете современных представлений о механизме проводимости, существование обедненных слоев толщиной порядка 1000 Å и потенциальных барьеров на границе зерен является определяющим фактором нелинейности ВАХ. При этом проводимость в области токов утечки объясняется термоактивированными токами и полевым понижением барьера. К настоящему моменту известен ряд моделей механизма проводимости оксидно-цинковой керамики на активном участке ВАХ с высокой нелинейностью. В области больших токов, ВАХ варистора снова стремится к омической зависимости. Величина ограничивающего сопротивления определяется удельным сопротивлением объема зерен ZnO, которое при концентрации свободных носителей порядка $10^{17} \dots 10^{18} \text{ см}^{-3}$ не превышает 0,3 Ом·см.

ВАХ оксидно-цинкового варистора обычно изображают в двойном логарифмическом масштабе для охвата широкого диапазона токов и напряжений (рис. 45). В таком масштабе любой участок ВАХ с постоянной нелинейностью представляется прямой линией и обычно аппроксимируется по следующему выражению: $I = (U / U_1)^\alpha$, где U_1 — значение напряжения при единичном токе.

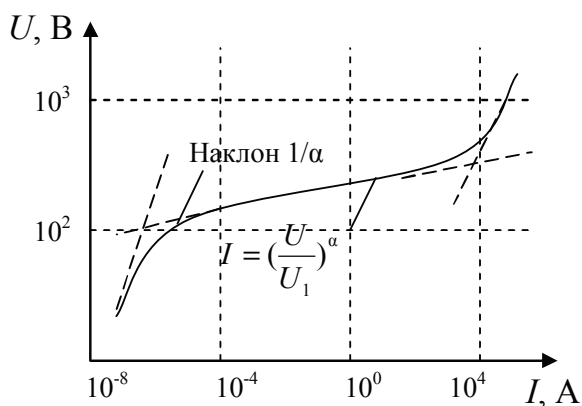


Рис. 45. ВАХ оксидного варистора

Разновидности варисторов

Промышленная маркировка — цифро-буквенное обозначение варисторов (например, СН1-2-2) — содержит информацию [7]:

- 1) СН — сопротивление нелинейное, т. е. варистор;
- 2) 1-я цифра означает материал (1 — карбид кремния; 2 — селен);
- 3) 2-я цифра (через дефис) — тип конструкции (1, 8 — стержневой; 2, 6, 7, 10 — дисковый; 3 — микромодульный);
- 4) 3-я цифра — порядковый номер разработки (как в нашем примере) или величину классификационного напряжения варистора (например, СН1-10-47).

Конструкции отечественных варисторов приведены на рис. 46.

Полное обозначение варистора, например, дискового типа с классификационным напряжением 270 В с допуском (± 10) %, такое: СН1-2-1-270 В ± 10 %.

Классификационное напряжение для стержневых варисторов определяется при кратковременном пропускании постоянного тока 10 мА, для дисковых — при токе 3 мА в диапазоне классификационных напряжений 15–27 В и при токе 2 мА в диапазоне 50–270 В.

Область применения варисторов

Исследования и разработки в области оксидно-полупроводниковых варисторов определены развитием полупроводниковой и микроэлектронной техники. Последняя, в отличие от электронно-лампо-

вой, крайне неустойчива к перенапряжениям, возникающим в реальной аппаратуре. Варисторы в таком случае эффективно используются для стабилизации напряжений, ограничения импульсных перегрузок, искрогашения на контактах, помехоподавления, управления люминесцентными индикаторами и жидкокристаллическими дисплеями, помехоподавления и искрогашения в коллекторных микро-двигателях, защиты спецаппаратуры от фактора электромагнитного импульса и т. д.

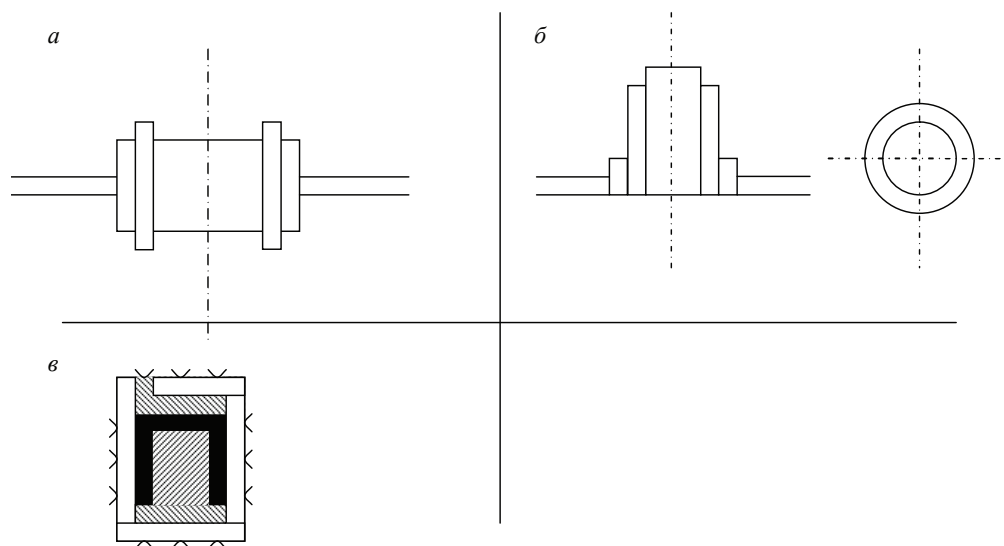


Рис. 46. Конструкции отечественных варисторов:

a — СН1; *б* — СН1-2; *в* — СН1-3

Область применения варисторов включает: системы телефонии и связи, системы железнодорожной связи и автоматики, тиристорные источники питания и преобразователи, вторичные источники питания, телевизионную технику, ряд областей электроники, бортовые радиоэлектронные системы, робототехнику, микроэлектронные системы, дискретные полупроводниковые приборы. В последнее время область их применения значительно расширилась: варисторы используются для ограничения перегрузок и помехоподавления в бытовой звуковой аппаратуре высокого класса, автомобильной микроэлектронике. Например, для защиты оборудования от импульсных напряжений в разных странах применяются вентильные разряд-

ники, RC -цепочки, LC -фильтры и др. Однако во всем мире наиболее эффективным (и дешевым) средством защиты от импульсных напряжений любого вида признано использование нелинейных полупроводниковых резисторов — варисторов.

Основной принцип действия варистора весьма прост. Варистор включается параллельно защищаемому оборудованию, т. е. при нормальной эксплуатации он находится под действием рабочего напряжения защищаемого устройства. В рабочем режиме (при отсутствии импульсных напряжений) ток через варистор пренебрежимо мал, и поэтому варистор в этих условиях представляет собой изолятор. При возникновении импульса напряжения, варистор, в силу нелинейности своей характеристики, резко уменьшает свое сопротивление до долей ома и шунтирует нагрузку, защищая ее и рассеивая поглощенную энергию в виде тепла. В этом случае через варистор может кратковременно протекать ток, достигающий нескольких тысяч ампер. Поскольку варистор практически безынерционен, то после гашения импульса напряжения он вновь приобретает очень большое сопротивление.

Таким образом, включение варистора параллельно электрооборудованию не влияет на его работу в нормальных условиях, но «срезает» импульсы опасного напряжения, что полностью обеспечивает сохранность даже ослабленной изоляции.

Наиболее широкое применение находят варисторы на основе оксида цинка, что обусловлено, во-первых, относительной простотой их изготовления и, во-вторых, хорошей способностью оксида цинка поглощать высокоэнергетические импульсы напряжения.

Стоит также отметить, что варисторы изготавливаются из функциональной керамики. Важным достоинством керамики является высокая доступность сырья, в том числе для получения бескислородной керамики типа карбидов и нитридов кремния, циркония или алюминия, заменяющих дефицитные металлы. Использование керамики открывает возможность для создания разнообразных по свойствам материалов в пределах одной и той же химической композиции. Любое, даже самое малое, керамическое изделие состоит из огромного числа кристаллитов, размером, формой и относительным расположением которых определяются их свойства. Отсюда возникает перспектива дальнейшей микроминиатюризации приборов с использованием керамических элементов.

7.2. ТЕРМОРЕЗИСТОРЫ

Терморезистором (от греч. *therme* — тепло, жар; лат. *resisto* — сопротивляюсь) обычно называют полупроводниковый резистор, электрическое сопротивление которого существенно убывает или возрастает с ростом температуры [7]. Для терморезистора характерны большой температурный коэффициент сопротивления (ТКС) (в десятки раз превышающий этот коэффициент у металлов), простота устройства, способность работать в различных климатических условиях при значительных механических нагрузках, стабильность характеристик во времени.

Разработаны следующие разновидности терморезисторов:

- 1) термистор — терморезистор, сопротивление которого с ростом температуры уменьшается. Термистор был открыт Самюэлем Рубеном (*Samuel Ruben*) в 1930 г. (патент США номер #2,021,491);
- 2) позистор — терморезистор, сопротивление которого с ростом температуры очень сильно возрастает;
- 3) терморезистор прямого подогрева, температура и сопротивление которого определяются температурой окружающей среды и саморазогревом от протекающего через него тока;
- 4) терморезистор косвенного подогрева — разогревается от специального дополнительного встроенного нагревателя;
- 5) болометр — терморезистор, чувствительный к воздействию теплового и оптического излучения, содержащий в своем составе активную и компенсационную части.

Широко известны температурные зависимости электропроводности металлов, собственных и примесных полупроводников (германий, кремний и др.). Однако в терморезисторах эти материалы не нашли применения по причине:

- 1) недостаточно сильной зависимости подвижности носителей заряда от температуры в проводниках и примесных полупроводниках;
- 2) несоответствия типовому диапазону рабочих температур $-60... +60^{\circ}\text{C}$ областей экспоненциального изменения концентрации носителей в примесных полупроводниках (область ионизации примесей менее 100 К, область перехода к собственной проводимости более 400 К);

- 3) высокой нестабильности величин сопротивлений технически изготавливаемых собственных полупроводников.

Принцип действия терморезисторов

Температурная зависимость сопротивления является главной характеристикой терморезисторов, в значительной степени определяющей остальные характеристики этих изделий. Она аналогична температурной зависимости удельного сопротивления полупроводника, из которого изготовлен данный терморезистор.

Измерения показывают, что температурная зависимость сопротивления большинства типов отечественных терморезисторов с отрицательным ТКС с достаточной для практики точностью во всем рабочем интервале температур или в его части аппроксимируется по выражению

$$R_T = A \exp (B/T),$$

где R_T — величина сопротивления терморезистора при температуре T , К; постоянная A зависит от физических свойств материала и габаритов терморезистора, $A = A_0 l/S$ (l — расстояние между электронами, см, и S — площадь поперечного сечения полупроводникового элемента терморезистора, см); постоянная B зависит от физических свойств материала и может иметь в интервале рабочих температур значения $B \approx \Delta W_3 / k$ при неполной ионизации примесей, $B \approx \Delta W_3 / 2k$ при собственной электропроводности.

Прологарифмировав аппроксимационное выражение, получим $\lg R_T = \lg A + 0,4343 B/T$. Это выражение в координатах R_T и $1/T$ представляет уравнение прямой, что значительно облегчает определение интервала температур, в котором формула с необходимой точностью аппроксимирует действительную зависимость $R_T(T)$. По результатам измерений R_T и T строят график зависимости $\lg R_T = f(1/T)$. Если через полученные экспериментально точки можно провести прямую, то считают, что в данном интервале температур выражение для R_T справедливо.

Для практических расчетов удобно исключить постоянную A . Написав формулу для RT для двух температур T_1 и T_2 и разделив одно на другое, получим

$$R_{T_2} = R_{T_1} \exp (B (T_1 - T_2)/T_1 T_2).$$

Из этой формулы можно рассчитать величину сопротивления терморезистора при любой температуре T_2 (в интервале рабочих температур), зная значение постоянной B и сопротивление образца при какой-то температуре T_1 .

Величина B определяется экспериментально путем измерения сопротивления терморезистора при температурах T_1 и T_2 . Логарифмируя предыдущее выражение, легко получить

$$B = (2,303 \cdot \Delta \lg R) / \Delta (1/T),$$

где B — коэффициент температурной чувствительности, К; $\Delta \lg R = \lg RT_2 - \lg RT_1$, $\Delta(1/T) = 1/T_2 - 1/T_1$.

Если определить ТКС терморезистора α , как это обычно принято $\text{ТКС} = \alpha_T = (1/R) (dR/dT)$, то следует, что $\alpha_T = -B/T^2$.

Для позисторов температурные зависимости сопротивления, снятые в широких интервалах температур, имеют сложный характер. При достаточно низких и высоких температурах, сопротивление уменьшается при увеличении температуры по закону близкому к экспоненциальному. В промежуточной области сопротивление R резко возрастает при повышении температуры. Крутизной графика, а следовательно, и величиной ТКС, можно управлять в широких пределах различными технологическими приемами.

Для многих типов позисторов, сопротивление в довольно большом интервале температур (порядка нескольких десятков градусов Цельсия или кельвина) меняется строго по экспоненциальному закону

$$R_T = A \exp(\alpha T),$$

где A — постоянная; α — температурный коэффициент сопротивления при температуре 1 °С в абсолютных единицах.

Принцип действия термисторов

В термисторах прямого подогрева, сопротивление изменяется или под влиянием теплоты, выделяющейся в них при прохождении электрического тока, или в результате изменения температуры термистора из-за изменения его теплового облучения (например, при изменении температуры окружающей среды).

Уменьшение сопротивления полупроводника с увеличением температуры (отрицательный ТКС) может быть вызвано разными причинами: увеличением концентрации носителей заряда, увеличением

интенсивности обмена электронами между ионами с переменной валентностью или фазовыми превращениями полупроводникового материала.

Увеличение концентрации носителей заряда характерно для термисторов, изготовленных из монокристаллов ковалентных полупроводников (кремний, германий, карбид кремния, соединения типа $A^{III}B^V$ и др.). Такие полупроводники обладают отрицательным температурным коэффициентом сопротивления в диапазоне температур, соответствующих примесной электропроводности, когда не все примеси ионизированы, а также в диапазоне температур собственной электропроводности, когда концентрация носителей изменяется из-за ионизации собственных атомов полупроводника. И в том и в другом случае зависимость сопротивления полупроводника определяется в основном по изменению концентрации носителей заряда, т. к. температурные изменения подвижности при этом пренебрежимо малы.

В этих диапазонах температур собственной и примесной электропроводности зависимость сопротивления полупроводника от температуры определяется уравнением

$$R_T = A \exp(B/T).$$

При неполной ионизации примесей и отсутствии компенсации, $B \sim \Delta W_{\Pi}/(2k)$, где ΔW_{Π} — энергия ионизации примесей (доноров или акцепторов); k — постоянная Больцмана. При собственной электропроводности, $B \sim \Delta W_3/(2k)$, где ΔW_3 — ширина запрещенной зоны полупроводника.

Основная часть термисторов, выпускаемых промышленностью, изготовлена из оксидных полупроводников — оксидов металлов переходной группы таблицы Д. И. Менделеева (от титана до цинка). Такие термисторы в форме стержней, трубок, дисков или пластинок получают методом керамической технологии, т. е. путем обжига заготовок при высокой температуре.

Электропроводность оксидных полупроводников с преобладающей ионной связью отличается от электропроводности ковалентных полупроводников. Для металлов переходной группы характерно наличие незаполненных электронных оболочек и переменная валентность. В результате этого, при образовании оксида в определенных условиях (наличие примесей, отклонение от стехиометрии), в одинаковых кри-

сталлографических положениях оказываются ионы с разными зарядами. Электропроводность таких материалов связана с обменом электронами между соседними ионами. Энергия, необходимая для такого обмена, экспоненциально уменьшается с увеличением температуры. В результате изменения интенсивности обмена электронами между ионами, температурная зависимость сопротивления термистора из оксидного полупроводника имеет такой же характер, как у термисторов из ковалентных полупроводников, но коэффициент температурной чувствительности в этом случае отражает изменение интенсивности обмена электронами между ионами, а не изменение концентрации носителей заряда.

В оксиде ванадия V_2O_4 и V_2O_3 при температуре фазовых превращений ($+68^\circ\text{C}$ и -110°C) наблюдается уменьшение удельного сопротивления на несколько порядков. Это явление также может быть использовано для создания термисторов с большим отрицательным температурным коэффициентом сопротивления в диапазоне температур, соответствующих фазовому превращению.

Основные используемые материалы для термисторов:

- 1) кобальт-марганцевые сплавы (например, терморезисторы КМТ-4);
- 2) медно-марганцевые сплавы (например, ММТ-1);
- 3) медно-кобальтовые сплавы (например, М2К2);
- 4) оксиды ванадия V_2O_4 и V_2O_3 .

Термистор косвенного подогрева — это термистор, имеющий дополнительный источник теплоты — подогреватель.

Конструктивное исполнение термисторов косвенного подогрева может быть разным. Часто подогреватель делают в виде обмотки на изоляционной трубке, внутри которой расположен термистор. В других случаях сам термистор сделан в виде трубки, внутри которой проходит нить подогрева. Общим для всех конструкций термисторов косвенного подогрева является наличие у них двух электрически изолированных друг от друга цепей: управляющей и управляемой.

Принцип действия болометров

Полупроводниковый болометр — это прибор, предназначенный для индикации и измерения теплового излучения (оптического или инфракрасного диапазона частот электромагнитного излучения).

Болометр состоит из двух пленочных термисторов (толщиной до 10 мкм). Один из термисторов болометра является активным, т. е. непосредственно подвергается воздействию измеряемого излучения. Сопротивление этого термистора изменяется в результате нагрева при облучении электромагнитным излучением оптического или инфракрасного диапазона частот. Другой термистор — компенсационный — служит для компенсации возможных изменений температуры окружающей среды. Компенсационный термистор должен быть экранирован от измеряемого излучения. Активный и компенсационный термисторы помещают в один герметичный корпус.

Болометры имеют три внешних вывода: от активного и компенсационного термисторов и от средней точки.

Полупроводниковые болометры применяют в различных системах ориентации, для бесконтактного и дистанционного измерения температур и др.

Принцип действия позисторов

Позистор — это полупроводниковый терморезистор с положительным температурным коэффициентом сопротивления [7].

В массовом производстве позисторы делают на основе керамики из титаната бария. Титанат бария BaTiO_3 — диэлектрик, с удельным сопротивлением при комнатной температуре $10^{10}–10^{12}$ Ом·см, которое значительно превышает удельное сопротивление полупроводников. Если же в состав керамики из титаната бария ввести примеси редкоземельных элементов (лантана, церия или др.) либо других элементов (ниобия, тантала, сурьмы, висмута и т. п.), имеющих валентность большую, чем у бария, и ионный радиус близкий к радиусу иона бария, то это приведет к уменьшению удельного сопротивления до $10–10^2$ Ом·см, которое соответствует удельному сопротивлению полупроводниковых материалов (рис. 47). Полупроводниковый титанат бария обладает аномальной температурной зависимостью удельного сопротивления: в узком диапазоне температур, при нагреве выше точки Кюри, удельное сопротивление полупроводникового титаната бария увеличивается на несколько порядков.

Механизм электропроводности полупроводникового титаната бария при наличии примесей можно представить следующим образом. Примесь редкоземельного элемента (например, лантана) замещает в узле

кристаллической решетки барий. Часть атомов титана, поддерживая электрическую нейтральность всего кристалла, захватывает лишние валентные электроны лантана, имеющего большую валентность, чем валентность бария. Захватываемые электроны, находясь в квазиустойчивом состоянии, легко перемещаются под действием электрического поля и обуславливают электропроводность материала

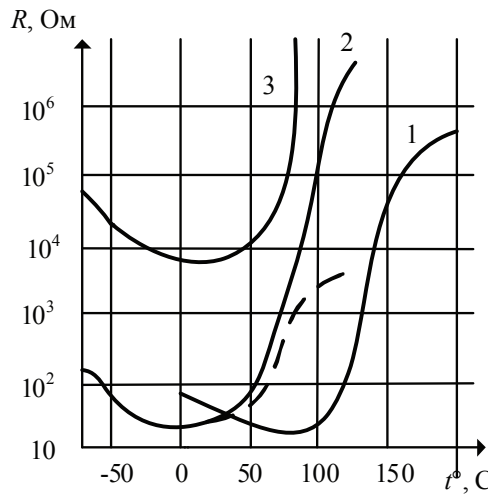
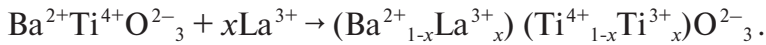
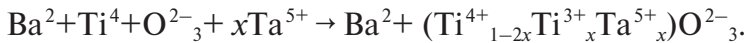


Рис. 47. Температурные характеристики некоторых позисторов:

1 — СТБ-1; 2 — СТБ-2Б; 3 — СТБ-3Б; штриховая кривая для СТБ-2Б получена при нагреве его проходящим током

При замещении титана в титанате бария примесями другого элемента (например, атомами тантала) аналогичный процесс электропроводности можно представить так:



Полупроводниковые свойства керамики на основе титаната бария можно получить также методом частичного восстановления



Во всех перечисленных случаях в полупроводниковом титанате бария существуют четырехвалентные и трехвалентные ионы титана.

Между разновалентными ионами титана может происходить обмен электронами. При этом каждый ион титана становится то трех-, то четырехвалентным (рис. 48). Данный процесс является причиной электропроводности полупроводникового титаната бария.

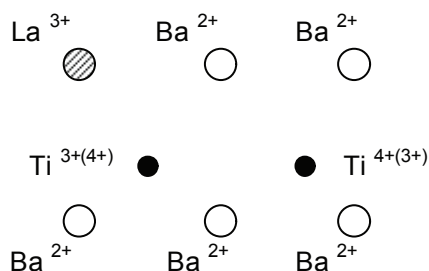


Рис. 48. Схема электропроводности титаната бария

Появление полупроводниковых свойств в ионных кристаллах под влиянием примесей наблюдается также и для оксида никеля. Полупроводники, изготавливаемые подобным методом, иногда называют полупроводниками с управляемой валентностью.

Технология изготовления позисторов аналогична технологии изготовления изделий из других керамических материалов. После смешивания исходных компонентов и веществ, содержащих примесные элементы, проводят их первичный обжиг при температуре около 1000 °С. Полученную твердую массу измельчают, а затем формуют заготовки (обычно в виде дисков различных размеров). Вторичный обжиг производят при температуре 1300–1400 °С.

Таким образом, резистивный слой позистора состоит из большого числа контактирующих между собой зерен или кристаллов полупроводникового титаната бария. Сопротивление позистора определяется сопротивлением обедненных поверхностных слоев на зернах титаната бария. Высота поверхностных потенциальных барьеров оказывается малой при температурах ниже точки Кюри, когда в зернах титаната бария существует спонтанная поляризация и материал обладает очень большой диэлектрической проницаемостью. При температурах больших точки Кюри, титанат бария претерпевает фазовое превращение из сегнетоэлектрического в параэлектрическое состояние. При этом пропадает спонтанная поляризация, резко уменьшается диэлектрическая проницаемость, растет высота поверхностных потен-

циальных барьеров на зернах титаната бария и увеличивается сопротивление позистора.

Температурные характеристики позисторов связаны с точкой Кюри керамики. Точка Кюри титаната бария может быть смещена в сторону низких температур путем частичного замещения бария стронцием. И наоборот, точка Кюри может быть смещена в сторону больших температур путем частичной замены бария свинцом. Уменьшает точку Кюри и частичная замена титана цирконием, оловом и самарием. Такое регулирование позволяет создавать позисторы, у которых положительный температурный коэффициент сопротивления наблюдается в разных диапазонах температур.

Титанат бария, конечно, не единственный сегнетоэлектрический материал, пригодный для изготовления позисторов. Положительный температурный коэффициент сопротивления можно получить также при использовании материалов системы SrTiO_3 с добавкой ниобия, системы PbTiO_3 с добавлением лантана, системы PbNbO_6 с добавкой вольфрама и др. Для совершенствования позисторов необходимы исследования материалов с точки зрения повторяемости свойств, надежности и увеличения срока службы, т. е. качеств, необходимых для внедрения в массовое производство.

Иногда для создания позисторов используют монокристаллические кремний, германий и другие полупроводниковые материалы. Принцип действия таких позисторов основан на уменьшении подвижности носителей заряда с увеличением температуры в результате увеличения их рассеяния на тепловых колебаниях атомов кристаллической решетки. Так, позисторы, изготовленные из монокристаллического кремния с малой концентрацией примесей ($10^{15} - 10^{17} \text{ см}^{-3}$), могут иметь температурный коэффициент сопротивления $(0,7 \dots 1) \cdot 10^{-2} \text{ K}^{-1}$ с положительным знаком в диапазоне температур $20 \dots \sim 100^\circ \text{C}$.

Одним из преимуществ монокристаллических позисторов перед поликристаллическими является возможность изготовления их с относительно малым разбросом параметров и характеристик. При использовании для изготовления монокристаллических позисторов с помощью широко распространенной планарной технологии можно создавать монокристаллические позисторы с разбросом номинальных сопротивлений в 1–2 %. Однако, из-за меньшей стоимости и из-за больших температурных коэффициентов сопротивления, поликристаллические позисторы нашли более широкое применение.

Параметры и характеристики терморезисторов

Основными параметрами терморезисторов являются следующие:

- 1) величина сопротивления образцов R_t и R_T , Ом, при определенной температуре окружающей среды t , °С, или T , К. Для терморезисторов, рассчитанных на рабочие температуры $-100...+125...200$ °С, температура окружающей среды принимается равной 20 или 25 °С. В таком случае величина R_t называется холодным сопротивлением, или номинальным. У различных терморезисторов номинальное сопротивление лежит в пределах от десятков омов до сотен килоомов, при этом допустимые отклонения от номинального сопротивления могут составлять $\pm (5, 10, 20) \%$;
- 2) коэффициент температурной чувствительности B , К,

$$B = \frac{T_1 T_2}{T_2 - T_1} \ln \frac{R_1}{R_2}.$$

У основной массы терморезисторов величина B лежит в диапазоне 2000–7200 К, но есть терморезисторы с $B = 700...15\,800$ К;

- 3) величина ТКС α , проценты на 1 °С. Обычно она указывается для той же температуры t , что и холодное сопротивление, и в этом случае обозначается через α_t

$$\alpha_t = \frac{dR}{RdT} \cdot 100\% = -\frac{B}{T^2}.$$

Значения ТКС при комнатной температуре для термисторов находятся в пределах 0,8–6 %/град., у позисторов — 10–20 %/град.;

- 4) постоянная времени τ , с. Она характеризует тепловую инерционность терморезистора и равна времени, в течение которого температура терморезистора изменяется на 63 % от разности температур образца и окружающей среды. Чаще всего эту разность берут равной 100 °С. Другими словами, τ — это промежуток времени, в течение которого температура терморезистора, перенесенного из спокойного воздуха $T = 0$ °С в спокойный воздух при $T = 100$ °С, достигнет температуры 63 °С (т. е. увеличится в $e = 2,72$ раза). Постоянная времени определяется в соответствии с конструкцией и размерами термистора, зависит от теплопроводности окружающей среды, составляет 0,5–140 с;

- 5) максимально допустимая температура t_{\max} , до которой характеристики терморезистора долгое время остаются стабильными;
- 6) максимально допустимая мощность рассеивания P_{\max} , Вт, не вызывающая необратимых изменений характеристик терморезистора. Естественно, при нагрузке терморезистора мощностью P_{\max} его температура не должна превышать t_{\max} ;
- 7) коэффициент рассеяния H , 1 Вт на 1 °С (К). Он равен мощности, рассеиваемой на терморезисторе при разности температур образца и окружающей среды в 1 °С, или, другими словами, равен мощности, которую надо выделить в термисторе, чтобы нагреть его на один градус;
- 8) коэффициент энергетической чувствительности G , Вт/%, равен мощности, которую нужно рассеять на терморезисторе для уменьшения его сопротивления на 1 %. Коэффициенты рассеяния и энергетической чувствительности зависят от параметров полупроводникового материала и от характера теплообмена между образцом и окружающей средой. Величины G , H и α связаны соотношением $G = \frac{H}{100\alpha}$. В самом деле,

$$G = \frac{\Delta P \cdot R}{\Delta R \cdot 100} = \frac{\frac{\Delta P}{\Delta T}}{\frac{100 \cdot \Delta R}{\Delta T \cdot R}} = \frac{H}{100\alpha};$$

- 9) теплоемкость C , 1 Дж на 1 °С, равна количеству тепла (энергии), необходимому для повышения температуры терморезистора на 1 °С. Можно доказать, что τ , H и C связаны между собой следующим соотношением:

$$\tau = \frac{C}{H}.$$

Для позисторов, кроме ряда приведенных выше параметров, обычно указывают в справочных данных также еще примерное положение интервала положительного температурного коэффициента сопротивления, а также кратность изменения сопротивления в области положительного ТКС.

Основными характеристиками терморезисторов являются следующие:

- 1) ВАХ — зависимость напряжения на терморезисторе от тока, проходящего через него. Снимается в условиях теплового равнове-

сия между теплотой, выделяемой в терморезисторе, и теплотой, отводимой от него в окружающую среду. Статическая ВАХ снимается в установившемся режиме с учетом постоянной времени терморезистора τ .

Начальные участки ВАХ термисторов и позисторов (OA , OC , OE на рис. 49) практически линейны. При дальнейшем увеличении тока, подводимая мощность возрастает, происходит саморазогрев терморезисторов и подводимое напряжение у термисторов (а, б) или незначительно возрастает (участок AB на рис. 49), или незначительно уменьшается (участок CD на рис. 49) из-за уменьшения их сопротивления.

У позисторов (в) в точке E происходит разогрев от подводимой мощности до температуры, соответствующей точке Кюри, и, при дальнейшем увеличении подводимого напряжения, ток резко уменьшается (участок EF), а сопротивление возрастает;

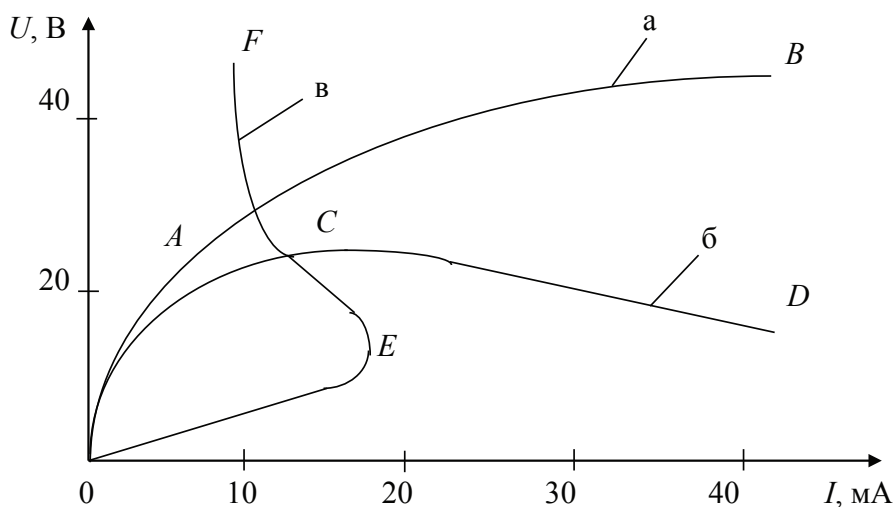


Рис. 49. Вольт-амперные характеристики терморезисторов:

a , b — термисторы ($TKC < 0$); v — позистор ($TKC > 0$)

- 2) температурная характеристика (рис. 50) — это зависимость $R(T)$, снимается в установившемся режиме;
- 3) подогревная характеристика (рис. 51) — характеристика, свойственная терморезисторам косвенного подогрева — зависимость сопротивления резистора от подводимой мощности.

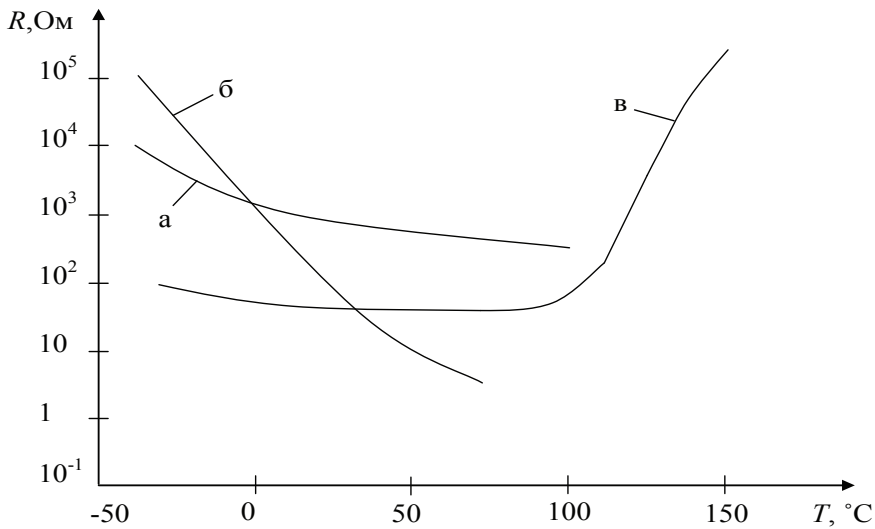


Рис. 50. Температурные характеристики терморезисторов:
а — термистор с $B = 2000$ К; б — термистор с $B = 5000$ К; в — позистор

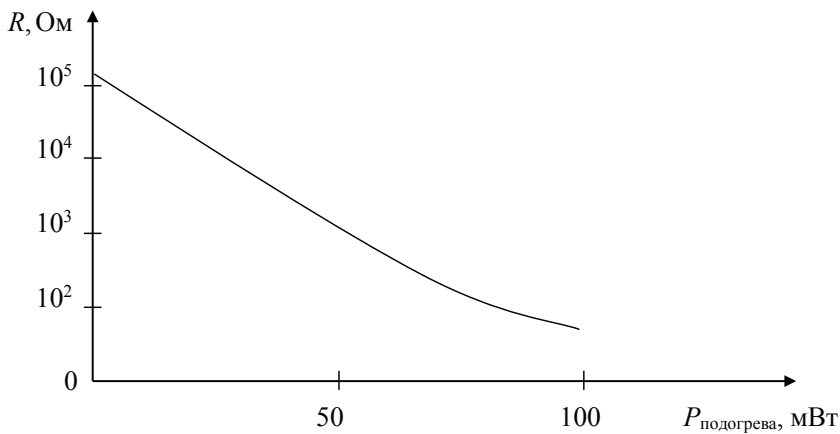


Рис. 51. Подогревная характеристика термистора косвенного подогрева

Разновидности и маркировка терморезисторов

Наиболее распространенные терморезисторы изготавливают на основе медно-марганцевых (ММТ и СТЗ), кобальтомарганцевых (КМТ и СТ1) и медно-кобальтомарганцевых (СТЗ) оксидных полупроводников [7].

По конструктивному оформлению терморезисторы можно разделить на следующие типы:

- 1) в виде цилиндрических стержней (КМТ-1, ММТ-1, КМТ-4, ММТ-4);
- 2) дисков (СТ1-17, СТ3-17, СТ5-1);
- 3) миниатюрных бусинок (СТ1-18, СТ1-19 и др.);
- 4) плоских прямоугольников (СТ3-23).

Особенностью бусиновых терморезисторов типов СТ1-18, СТ3-18 и СТ3-25 является то, что термочувствительный элемент для защиты от внешних воздействий покрыт тонким слоем стекла, а тонкие платиновые контакты приварены или припаяны (СТ3-25) к траверсам из толстой проволоки.

Терморезисторы ММТ-1 и КМТ-1 предназначены для работы в закрытых сухих помещениях, ММТ-4 и КМТ-4 герметизированы, работоспособны в условиях с повышенной влажностью и даже в жидкой среде.

Существуют также измерительные терморезисторы, предназначенные для измерений в маломощных цепях сверхвысокочастотных колебаний, — терморезисторы типа ТП (ТП2/0,5, ТП2/2, ТП6/2 — цифра в числителе — номинальное значение напряжения в вольтах, знаменатель — рабочий ток в миллиамперах) для стабилизации напряжения в цепях постоянного или переменного тока с частотой до 150 кГц; по конструкции — круглые опрессованные стержни, заключенные в стеклянный баллон, воздух из которого откачан до давления 10^{-5} мм рт. ст.

Терморезисторы ТКП, СТ1-21, СТ3-21 и СТ3-27 применяются в радиотехнических устройствах и схемах автоматики как регулируемые бесконтактные резисторы. Они имеют косвенный подогрев от специальной спирали, при изменении тока в которой происходит плавное изменение сопротивления терморезистора. Данные терморезисторы используются, когда необходимо отделить управляемую цепь от управляющей. Рабочий элемент и подогреватель терморезисторов типа ТКП помещен в стеклянный баллон с нормальным октальным цоколем. Терморезисторы типов СТ1-21, СТ3-21 и СТ3-27 (более новые) имеют более совершенную конструкцию по сравнению с ТКП.

Область применения

Терморезисторы широко применяются везде: на них основаны системы противопожарной безопасности, системы измерения и регулирова-

ния температуры, теплового контроля, схемы температурной компенсации, измерения мощности ВЧ и т. д. Также терморезисторы находят применение в промышленной электронике и бытовой аппаратуре, медицине, метеорологии, химической и других отраслях промышленности.

Однако применение терморезисторов в качестве датчиков температуры имеет не только плюсы, но и минусы. Так, например, это инерционность, обусловленная постоянной времени τ , плохая стабильность в определенных условиях и др.

7.3. ФОТОРЕЗИСТОРЫ

Фоторезистор — это полупроводниковый резистор, сопротивление которого изменяется в зависимости от поглощаемого светового потока, т. е. это полупроводниковый резистор, действие которого основано на фоторезистивном эффекте [7].

Принцип действия фоторезисторов

Фотопроводимость — это свойство полупроводника изменять свою электропроводность под воздействием электромагнитного излучения. Причина фотопроводимости — увеличение концентрации носителей заряда — электронов — в зоне проводимости и дырок в валентной зоне. Вследствие этого проводимость полупроводника возрастает на величину

$$\Delta\sigma = q (\mu_n \Delta n_i + \mu_p \Delta p_i),$$

где q — заряд электрона; μ_n — подвижность электронов; μ_p — подвижность дырок; Δn_i — концентрация генерируемых электронов; Δp_i — концентрация генерируемых дырок.

Поскольку основным следствием поглощения энергии света в полупроводнике является перевод электронов из валентной зоны в зону проводимости, энергия кванта света фотона должна удовлетворять условию

$$h\nu_{кр} \geq \Delta W_3 \text{ или } h\nu_{кр} \geq \Delta W_{пр},$$

где h — постоянная Планка; $\nu_{кр}$ — критическая частота электромагнитного излучения (красная граница фотопроводимости); ΔW_3 — ширина запрещенной зоны; $\Delta W_{пр}$ — энергия ионизации примеси.

Излучение с частотой $\nu < \nu_{\text{кр}}$ не может вызвать фотопроводимость, т. к. энергия кванта такого излучения $h\nu < \Delta W_3$ недостаточна для перевода электрона из валентной зоны в зону проводимости. Если же $h\nu > \Delta W_3$, то избыточная относительно ширины запрещенной зоны часть энергии квантов передается электронам в виде кинетической энергии.

Критической частоте $\nu_{\text{кр}}$ соответствует граничная длина волны

$$\lambda_{\text{гр}} = c/\nu_{\text{кр}},$$

где c — скорость света.

При длинах волн больших граничной, фотопроводимость резко падает. Так, для германия граничная длина волны составляет примерно 1,8 мкм. Однако спад фотопроводимости наблюдается и в области малых длин волн. Это объясняется быстрым увеличением поглощения энергии с частотой и уменьшением глубины проникновения падающей на полупроводник электромагнитной энергии. Поглощение происходит в тонком поверхностном слое, где и образуется основное количество носителей заряда. Появление большого количества избыточных носителей только у поверхности слабо отражается на проводимости всего объема полупроводника, т. к. скорость поверхностной рекомбинации больше объемной и проникающие вглубь неосновные носители заряда увеличивают скорость рекомбинации в объеме полупроводника.

Фотопроводимость полупроводников может обнаруживаться в инфракрасной, видимой или ультрафиолетовой частях электромагнитного спектра в зависимости от ширины запрещенной зоны, которая зависит от типа полупроводника, температуры, концентрации примесей и напряженности электрического поля.

Рассмотренный механизм поглощения света, приводящий к появлению свободных носителей заряда в полупроводнике, называют фотоактивным. Поскольку изменяется проводимость, а следовательно, внутреннее сопротивление полупроводника, указанное явление было названо фоторезистивным эффектом.

При облучении фоторезистора фотонами, в полупроводниковом фоточувствительном слое возникает избыточная концентрация носителей заряда. Если к фоторезистору приложено напряжение, то через него будет проходить дополнительная составляющая тока — фототок, обусловленный избыточной концентрацией носителей. Электронная составляющая фототока

$$I_{\phi n} = abq\Delta n\mu_n \frac{U}{l} = abq(1-R)\alpha\eta N_{\phi}\tau_n\mu_n E,$$

где a — толщина полупроводникового фоточувствительного слоя; b — его ширина; l — расстояние между электродами; R — коэффициент отражения; α — показатель поглощения; η — квантовая эффективность генерации; N_{ϕ} — число фотонов, падающих на единичную поверхность фоточувствительного слоя за единицу времени.

Фототок соответствует прохождению через фоторезистор и через внешнюю цепь $I_{\phi n}/q$ электронов. Число электронов, возникающих в объеме фоточувствительного слоя из-за поглощения фотонов, равно $abl(1-R)\alpha\eta N_{\phi}$.

Отношение числа прошедших во внешней цепи электронов к числу возникших в фоточувствительном слое электронов называют коэффициентом усиления фоторезистора

$$K_I = \frac{I_{\phi n} / q}{abl(1-R)\alpha\eta N_{\phi}} = \frac{ab(1-R)\alpha\eta N_{\phi}\tau_n\mu_n E}{abl(1-R)\alpha\eta N_{\phi}} = \frac{\tau_n\mu_n E}{l}.$$

Произведение подвижности электронов на напряженность электрического поля есть скорость дрейфа электронов, которую можно также представить как расстояние между электродами, деленное на время пролета носителей между электродами $t_{\text{прол}}$. Поэтому коэффициент усиления фоторезистора можно выразить и в таком виде: $K_I = \tau_n/t_{\text{прол}}$.

Если в полупроводниковом фоточувствительном слое есть примеси, являющиеся ловушками захвата для неосновных носителей заряда (сенсibiliзирующие примеси), то захват неосновных носителей этими ловушками может существенно (на несколько порядков) увеличить эффективное время жизни неравновесных основных носителей. В этом случае время жизни может значительно превышать время пролета носителей между электродами. Когда один из электродов достигает положительного электрода, другой электрон входит в полупроводниковый слой из отрицательного электрода для сохранения электрической нейтральности объема полупроводника, в котором осталась нескомпенсированная положительно заряженная ловушка захвата. Таким образом, поглощение одного фотона может служить причиной прохождения через фоторезистор многих электронов.

Введение сенсibiliзирующих примесей, приводя к увеличению эффективного времени жизни основных носителей, вызывает снижение быстродействия фоторезистора.

Усиление фототока может происходить и при наличии потенциальных барьеров, например, на поверхности кристаллов полупроводника, если фоторезистор изготовлен на основе поликристаллического полупроводникового материала. Потенциальные барьеры могут являться потенциальными ямами для неосновных носителей заряда. В этом случае будет происходить усиление фототока в фоторезисторе по аналогии с усилением фототока в фототранзисторе.

Параметры и характеристики фоторезисторов

Определим основные параметры фоторезисторов.

1. Фототок — ток, протекающий через фоторезистор при указанном напряжении на нем, обусловленный только воздействием потока излучения с заданным спектральным распределением.
2. Общий ток фоторезистора — ток, состоящий из темнового тока и фототока. Темновой ток измеряют при напряжении 1 В.
3. Рабочее напряжение U_p — постоянное напряжение, приложенное к фоторезистору, при котором обеспечиваются номинальные параметры при длительной его работе в заданных эксплуатационных условиях (гарантирующие продолжительную работу фоторезистора). При работе в импульсном режиме, у селенисто-кадмиевых и селенисто-кадмиевых фоторезисторов допустимое напряжение может в 2–3 раза превышать рабочее. У селенисто-свинцовых фоторезисторов рабочее напряжение можно принять равным $0,1 R_T$, где R_T в килоомах. У различных фоторезисторов U_p лежит в пределах 4–400 В.
4. Максимально допустимое напряжение фоторезистора U_{\max} — максимальное значение постоянного напряжения, приложенного к фоторезистору, при котором отклонение его параметров от номинальных значений не превышает указанных пределов при длительной работе в заданных эксплуатационных условиях.
5. Значения максимально допустимой мощности рассеяния P_{\max} находятся в пределах 0,01–0,2 Вт. Она ограничена допустимым ростом темнового тока за счет разогрева.
6. Темновое сопротивление R_T — сопротивление фоторезистора в отсутствие падающего на него излучения в диапазоне его спек-

тральной чувствительности. У некоторых типов фоторезисторов темновое сопротивление может иметь значительный разброс 10^4 – 10^7 Ом.

7. Световое сопротивление R_C — сопротивление фоторезистора, измеренное через определенный интервал времени после начала воздействия излучения, создающего на нем освещенность заданного значения.
8. Кратность изменения сопротивления K_R — отношение темнового сопротивления фоторезистора к сопротивлению при определенном уровне освещенности (световому сопротивлению) $K_R = R_T / R_C$. Это один из важнейших параметров, характеризующих чувствительность фоторезистора. С увеличением освещенности, кратность возрастает по линейному закону, с уменьшением — снижается. Наименьшей чувствительностью обладают сернисто-свинцовые фоторезисторы, у которых кратность при освещенности 200 лк не ниже 1,2. У остальных типов фоторезисторов чувствительность значительно выше и может быть 10^2 – 10^4 .
9. Постоянная времени τ_ϕ — время, в течение которого фототок изменяется на 63 %, т. е. в e раз. Постоянная времени характеризует инерционность прибора и влияет на вид его частотной характеристики.
10. Коэффициент усиления по току M может достигать значений 10^3 – 10^7 . Для увеличения M увеличивают постоянную времени τ (при этом возрастает инерционность и уменьшается диапазон частот модуляции сигнала) либо уменьшают время пролета (уменьшая длину фоточувствительного слоя и увеличивая рабочее напряжение).
10. Длинноволновая граница фотопроводимости λ_0 , мкм, определяется по ширине запрещенной зоны ΔW_3 либо энергии активации примесей $\Delta W_{пр}$, эВ, используемых материалов.

$$\lambda_0 = 1,236 / \Delta W.$$

Наиболее распространенные материалы фоторезисторов имеют следующие значения:

- сернистый кадмий (CdS) $\lambda_0 = 0,49$ мкм ($\Delta W = 2,53$ эВ);
- сернистый свинец (PbS) $\lambda_0 = 3,36$ мкм ($\Delta W = 0,37$ эВ);

- антимонид индия (InSb, 77 K) $\lambda_0 = 7,3$ мкм ($\Delta W = 0,17$ эВ);
 - германий, легированный медью (Ge: Cu, 15 K), $\lambda_0 = 25$ мкм.
12. Рабочая длина волны λ_{\max} соответствует максимуму спектральной характеристики фоторезистора, λ_{\max} на 20–50 % < λ_0 .
13. Температурный коэффициент фототока

$$\alpha_T = \frac{dI}{I} \frac{1}{dT} \cdot 100\% = -(0,1 \dots 0,01)\% / \text{град}.$$

С ростом температуры фототок уменьшается в основном из-за уменьшения времени жизни τ носителей и уменьшения коэффициента усиления M .

14. Пороговая (обнаружительная) чувствительность Φ_{\min} — минимальный световой поток, вызывающий появление фототока, равного темновому току. Она определяется в полосе $\Delta f = 1$ Гц при соотношении сигнал/шум равном 1, приводится к площади светочувствительного слоя 1 см²; имеет, например, у фоторезистора из CdS значение $2 \cdot 10^{-15}$ Вт/(см·Гц^{1/2}).
15. Удельная интегральная чувствительность — отношение фототока к произведению величины падающего на фоторезистор светового потока на приложенное к нему напряжение

$$K_0 = I_{\phi} / (\Phi U),$$

где I_{ϕ} — фототок равный разности токов, протекающих по фоторезистору в темноте и при определенной (200 лк) освещенности; Φ — падающий световой поток; U — напряжение, приложенное к фоторезистору.

Чувствительность называют интегральной, потому что измеряют ее при освещении фоторезистора светом сложного спектрального состава. Для различных фоторезисторов K_0 составляет 1–600 мА/(лм·В).

Важными характеристиками фоторезисторов являются вольт-амперная, световая, спектральная, частотная, переходная, шумовая.

16. ВАХ — вольт-амперная характеристика $I = f(U)|_{\Phi=\text{const}}$, характеризующая зависимость фототока (рис. 52) (при постоянном световом потоке Φ) или темнового тока от приложенного напряжения. Закон Ома нарушается в большинстве случаев только при высоких напряжениях на фоторезисторе. Эта характеристика линейна в довольно широких пределах. У некоторых типов фоторезисторов при напряжениях меньше рабочего наблюдается нелинейность.

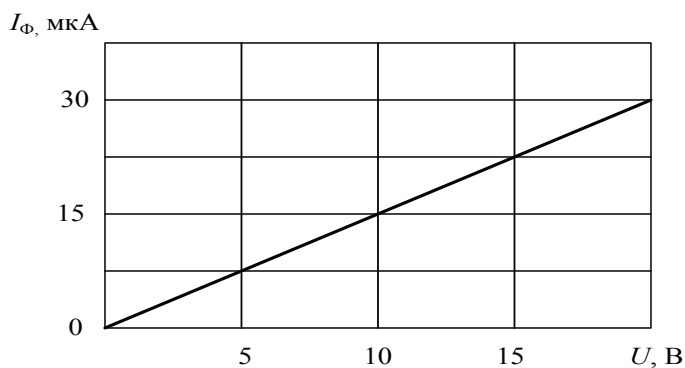


Рис. 52. Вольт-амперная характеристика фоторезистора для фототока при фиксированном световом потоке

Обычно рассматривают семейство ВАХ при различных значениях Φ (рис. 53).

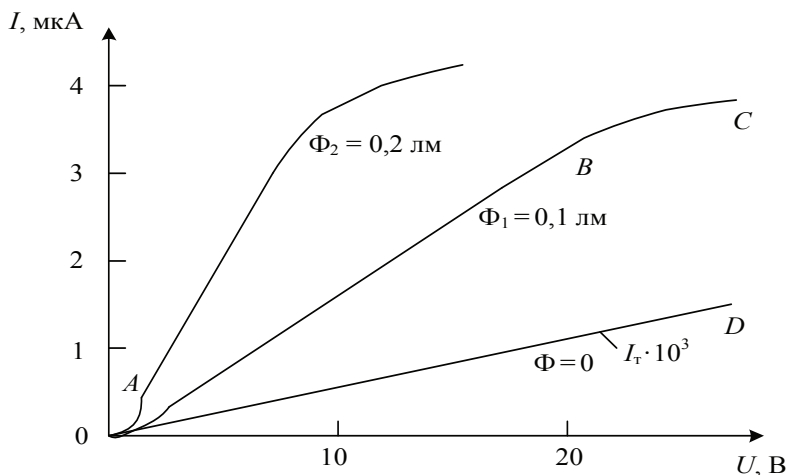


Рис. 53. Вольт-амперная характеристика фоторезистора

Темновая ВАХ фоторезистора (0D) линейна, для наглядности на рис. 53 значение темнового тока увеличено в 10^3 раз. Типовая ВАХ (например, при $\Phi_1 = 0,1 \text{ лм}$) в общем случае состоит из 3 участков:

- 0A — обычно нелинейный, из-за заметного при малых токах и напряжениях действия потенциальных барьеров между полупроводниковым слоем и выводами, между отдельными зернами фоточувствительного слоя (нарисован несколько укруп-

ненно). При малых напряжениях сопротивление фоторезистора определяется в основном сопротивлением контактов между отдельными зернами или кристаллами полупроводника, напряженность электрического поля на них получается большой. Поэтому, при увеличении приложенного напряжения, сопротивление контактов уменьшается либо из-за эффектов сильного поля (например, туннелирование сквозь узкие потенциальные барьеры на контактах), либо из-за разогрева приконтактных областей отдельных зерен полупроводника;

- AB — рабочий, линейный участок. В этом случае сопротивление фоторезистора постоянное и определяется объемным сопротивлением зерен полупроводника;
- BC — загиб ВАХ при больших токах — обусловлен разогревом фоточувствительного слоя, ростом концентрации носителей, скорости рекомбинации и уменьшением времени жизни носителей и коэффициента усиления M .

17. Световая (люкс-амперная) характеристика $I = f(\Phi)|_{U=\text{const}}$ характеризует зависимость фототока от падающего светового потока постоянного спектрального состава при постоянном напряжении на фоторезисторе. Полупроводниковые фоторезисторы имеют нелинейную люкс-амперную характеристику (рис. 54). Рабочей является линейная часть световой характеристики (участок AB), на которой ток фоторезистора пропорционален световому потоку.

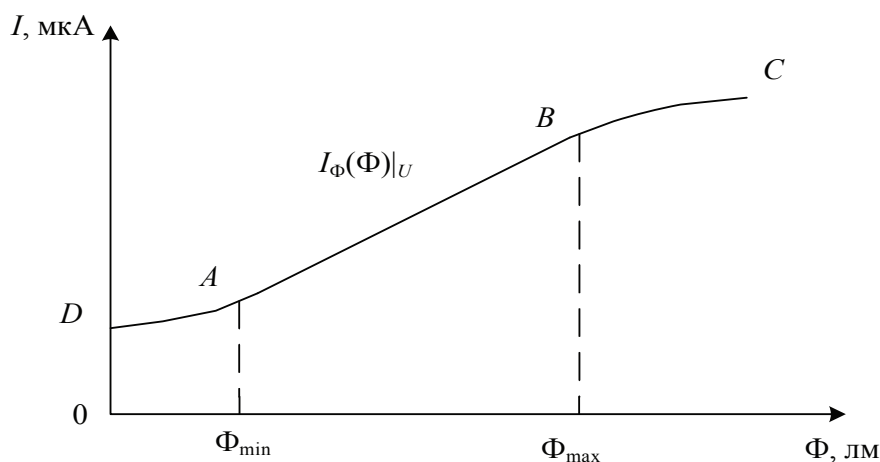


Рис. 54. Световая характеристика фоторезистора

При малых интенсивностях светового потока, световая характеристика (участок DA) нелинейна:

- из-за наличия темнового тока (при $\Phi = 0$);
- непропорционального роста фототока при слабом освещении в результате захвата фотоносителей центрами рекомбинации, ловушками, создаваемыми дефектами кристаллической решетки, атомами посторонних примесей. При увеличении освещенности происходит смещение квазиуровней Ферми для электронов и дырок к основным зонам (для дырок — к валентной зоне, для электронов — к зоне проводимости). В результате этого часть уровней ловушек захвата становится уровнями рекомбинационных потерь, что приводит к уменьшению времени жизни носителей заряда.

При больших световых потоках (участок BC) происходит загиб световой характеристики из-за увеличения скорости рекомбинации носителей, уменьшения времени жизни τ и коэффициента усиления M . Кроме того, уменьшается подвижность носителей заряда, т. к. при увеличении освещенности возрастает число ионизированных атомов в полупроводнике и, следовательно, растёт рассеяние носителей заряда ионизированными атомами.

Наибольшая чувствительность получается при малых освещенностях. Это позволяет использовать фоторезисторы для измерения очень малых интенсивностей излучения. При увеличении освещенности, световой ток растёт примерно пропорционально корню квадратному из освещенности. Наклон люкс-амперной характеристики зависит от приложенного к фоторезистору напряжения.

18. Спектральная характеристика — это зависимость относительной (т. е. нормированной по максимуму) чувствительности (тока) фоторезистора от длины волны излучения при фиксированных напряжении на фоторезисторе и силе светового потока (рис. 55). Спектральная характеристика определяет чувствительность фоторезистора.

При $\lambda > \lambda_0$ энергии кванта недостаточно для генерации свободных носителей заряда ($h\nu < \Delta W$).

На участке $\lambda_0 - \lambda_{\max}$ рост тока обусловлен ростом плотности разрешенных состояний в валентной зоне и зоне проводимости при удалении от их границ и возможностью поглощения потока квантов света.

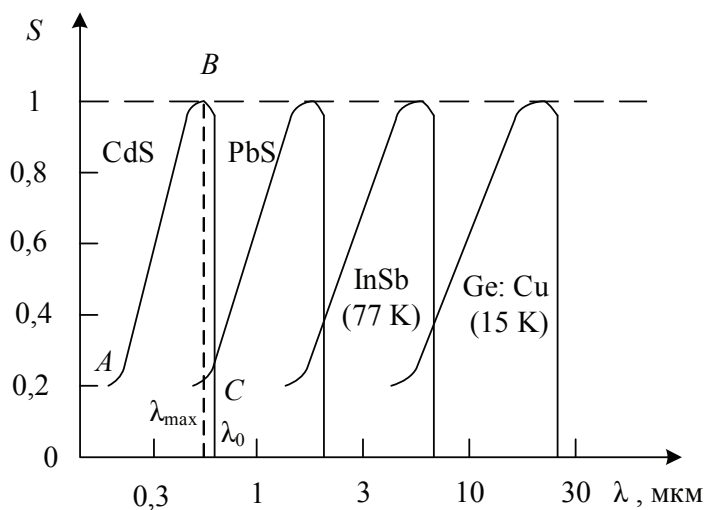


Рис. 55. Спектральные характеристики фоторезистора

Спад тока на участке *BA* при уменьшении λ обусловлен:

- уменьшением числа квантов света при постоянной силе светового потока Φ и уменьшением числа фотогенерированных носителей заряда;
- поглощением света в приповерхностном слое все меньшей толщины, т. е. уменьшением глубины проникновения квантов света в полупроводник и возрастанием потерь от поверхностной рекомбинации, уменьшением эффективного времени жизни фотогенерированных носителей заряда.

Спектральная характеристика определяется материалом, используемым для изготовления светочувствительного элемента. Как видно из характеристик, приведенных на рис. 55, фоторезисторы с селенисто-кадмиевым светочувствительным элементом имеют максимальную чувствительность в видимой части спектра, фоторезисторы, выполненные на основе селенистого кадмия, наиболее чувствительны к красной и инфракрасной части спектра, а селенисто-свинцовые фоторезисторы имеют максимум чувствительности в инфракрасной области спектра.

19. Частотная характеристика — это зависимость относительного значения фототока от частоты модуляции светового потока. Эта характеристика определяет чувствительность фоторезистора при действии на него светового потока, изменяющегося с опре-

деленной частотой. Наличие инерционности у фоторезисторов приводит к тому, что величина их фототока зависит от частоты модуляции падающего на них светового потока: с увеличением частоты светового потока фототок уменьшается (рис. 56). Инерционность ограничивает возможности применения фоторезисторов при работе с переменными световыми потоками высокой частоты.

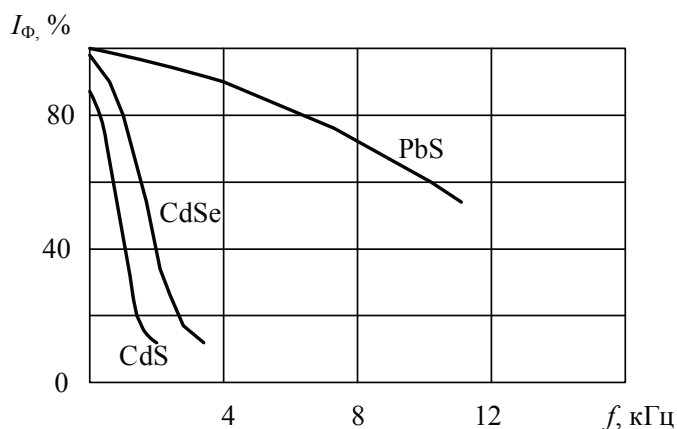


Рис. 56. Частотные характеристики фоторезистора

Как следует из анализа данной характеристики, величина сигнала, снимаемого с фоторезистора, уменьшается с увеличением частоты модуляции светового потока. Чувствительность фоторезисторов меняется (уменьшается) в первые 50 ч работы, оставаясь в дальнейшем практически постоянной в течение всего срока службы, измеряемого несколькими тысячами часов. Интервал рабочих температур для сернисто-кадмиевых фоторезисторов составляет $-60...+85\text{ }^{\circ}\text{C}$ для селенисто-кадмиевых — $-60...+40\text{ }^{\circ}\text{C}$ и для сернисто-свинцовых — $-60...+70\text{ }^{\circ}\text{C}$.

Разновидности и маркировка фоторезисторов

Полупроводниковые фоторезисторы работают в цепях как постоянного, так и переменного тока. Техническими условиями допускается также использование фоторезисторов в импульсных режимах при средней мощности рассеяния, не превышающей максимально допустимого значения. Фоторезисторы могут работать при больших ин-

тенсивностях света при условии непревышения предельного значения фототока и мощности рассеяния [7].

Для изготовления серийных фоторезисторов используются различные типы материалов: сернистый кадмий (CdS), селенистый кадмий (CdSe), сернистый свинец (PbS) и селенид свинца (PbSe). Фоторезисторы, чувствительные к инфракрасному излучению длинноволнового диапазона, изготавливают на основе соединения кадмий-ртуть-теллур и антимонида индия (InSb).

Светочувствительный элемент в некоторых типах фоторезисторов выполнен в виде круглой или прямоугольной таблетки, спрессованной из порошкообразного сульфида или селенида кадмия, в других — он представляет собой тонкий слой полупроводника, нанесенного на стеклянное основание. В том и в другом случае с полупроводниковым материалом соединены два металлических вывода. В зависимости от назначения фоторезисторы имеют различное конструктивное оформление. Иногда это просто пластина полупроводника на стеклянном основании с токонесущими выводами, в других случаях фоторезистор имеет пластмассовый корпус с жесткими штырьками. Среди таких фоторезисторов следует особо отметить ФСК-6, приспособленный для работы от отраженного света, для чего его корпус имеет в центре отверстие для прохождения света к отражающей поверхности. Выпускаются фоторезисторы в металлическом корпусе с цоколем, напоминающим ламповый, или в корпусе, как у герметизированных конденсаторов или транзисторов. Малогабаритные пленочные фоторезисторы выпускаются в пластмассовых и металлических корпусах с влагозащитным покрытием светочувствительного элемента прозрачными эпоксидными смолами.

Конструкция монокристаллического и пленочного фоторезисторов показана на рис. 57, 58. Основным элементом фоторезистора является в первом случае монокристалл, а во втором — тонкая пленка полупроводникового материала.

Если фоторезистор включен последовательно с источником напряжения (рис. 59) и не освещен, то в его цепи будет протекать темновой ток

$$I_T = E / (R_T + R_H),$$

где E — ЭДС источника питания; R_T — величина электрического сопротивления фоторезистора в темноте, называемая темновым сопротивлением; R_H — сопротивление нагрузки.

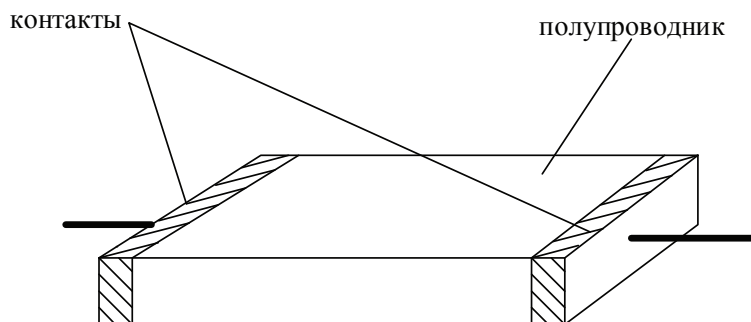


Рис. 57. Монокристаллический фоторезистор

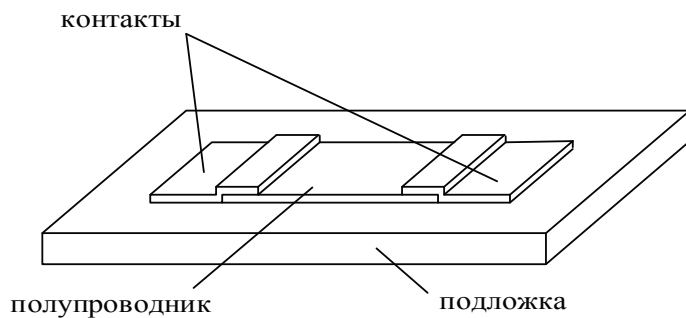


Рис. 58. Пленочный фоторезистор

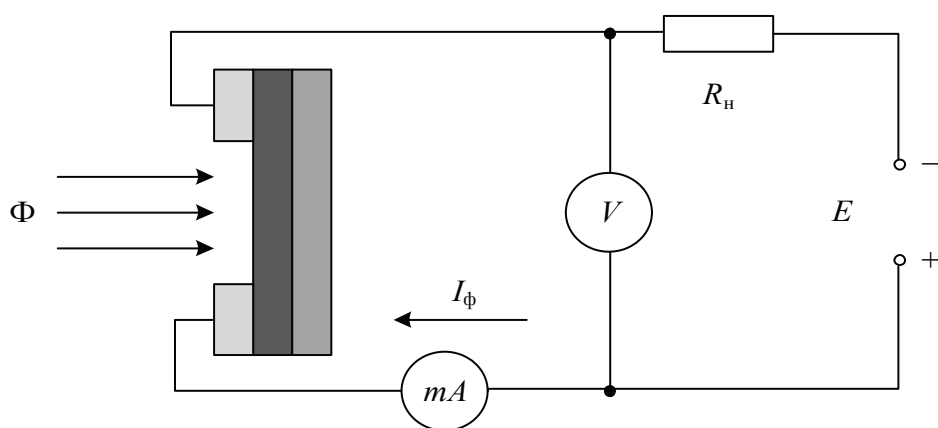


Рис. 59. Схема включения для измерения параметров и характеристик фоторезисторов

При освещении фоторезистора, энергия фотонов расходуется на перевод электронов в зону проводимости. Количество свободных электронно-дырочных пар возрастает, сопротивление фоторезистора падает, и через него течет световой ток

$$I_C = E / (R_C + R_H).$$

Разность между световым и темновым током дает значение тока I_Φ , получившего название первичного фототока проводимости $I_\Phi = I_C - I_T$.

Когда лучистый поток мал, первичный фототок проводимости практически безынерционен и изменяется прямо пропорционально величине лучистого потока, падающего на фоторезистор. По мере возрастания величины лучистого потока увеличивается число электронов проводимости. Двигаясь внутри вещества, электроны сталкиваются с атомами, ионизируют их и создают дополнительный поток электрических зарядов, получивший название вторичного фототока проводимости. Увеличение числа ионизированных атомов тормозит движение электронов проводимости. В результате этого изменение фототока запаздывает во времени по сравнению с изменением светового потока, что определяет некоторую инерционность фоторезистора.

Старое обозначение фоторезисторов состоит:

- из двух букв ФС;
- последующей буквы, указывающей на материал фоточувствительного элемента (К — CdS; Д — CdSe; А — PbS);
- цифры — указания на конструктивное оформление фоторезистора; иногда перед цифрой стоит буква Г или П, характеризующая конструктивное исполнение и означающая, что конструкция герметизирована и предназначена для условий тропического климата и повышенной влажности (Г — герметизированные, П — пленочные).

Современное обозначение фоторезисторов состоит из букв ФР и номера разработки. Например, ФР-193 Д (фоторезистор с номером разработки 193 группы Д). Название типа фоторезисторов складывается из букв и цифр, причем цифры обозначают тип использованного светочувствительного материала.

Ниже приведены наименования наиболее распространенных обозначений фоторезисторов:

ФСА-0, ФСА-1, ФСА-6, ФСА-Г1, ФСА-Г2..... сернисто-свинцовые
 ФСК-0,1,2,4,5,6,7, ФСК-Г1, ФСК-Г2,
 ФР-Г7, ФСК-П1 СФ2-1, 2, 4, 9, 12..... сернисто-кадмиевые
 ФСД-0, ФСД-1, ФСД-Г1 СФ3-1, 8..... селенисто-кадмиевые

Область применения

Благодаря высокой чувствительности, простоте и разнообразию конструкций, фоторезисторы имеют широкое применение в радиоэлектронике. Основными областями их применения являются: системы фотоэлектрической автоматики и телемеханики, измерительные приборы, экспонетрические приборы и фотодатчики, системы телеконтроля, а также тепловизионная аппаратура промышленного, медицинского и оборонного назначения.

Фоторезисторы, как элементы цепи преобразования информации, применяются в различных системах, предназначенных для контроля и изменения геометрических размеров и скоростей движения объектов, температуры, управления различными механизмами, для определения качественного и количественного состава твердых, жидких и газообразных сред, включения и выключения различных устройств и др.

Во многих случаях фоторезисторный способ получения и обработки информации дает явное преимущество по сравнению с другими способами. При их применении достигается односторонность связи между источником сигнала информации — излучателем и потребителем — фоторезистором.

Успехи в развитии фоторезисторной автоматики стали возможными благодаря значительному усовершенствованию конструкции и расширению номенклатуры серийно выпускаемых фоторезисторов. Они обладают высокой чувствительностью, достаточно малой инертностью, имеют незначительные габариты, долговечны в работе, обеспечивают бесконтактные изменения и контроль. Фоторезисторы с успехом заменяют вакуумные и газонаполненные фотоэлементы. Обладая повышенной допустимой мощностью рассеивания по сравнению с некоторыми типами фотоэлементов, фоторезисторы позволяют создавать простые и надежные фотореле без усилителей тока. Такие фотореле незаменимы в устройствах для телеуправления, контроля и регулирования, в автоматах для разбраковки, при сортировке и счете готовой продукции, для контроля качества и готовности самых различных деталей.

Широко используются фоторезисторы в полиграфической промышленности при обнаружении обрывов бумажной ленты, контроле за количеством листов, подаваемых в печатную машину.

В измерительной технике фоторезисторы применяются для измерения высоких температур, для регулировки температуры в различных технологических процессах.

Контроль уровня жидкости и сыпучих тел, защита от входа персонала в опасные зоны, контроль за запыленностью и задымленностью самых различных объектов, автоматические выключатели уличного освещения и турникеты в метрополитене — вот далеко не полный перечень областей применения фоторезисторов.

8. *p-n* ПЕРЕХОД

8.1. ОПРЕДЕЛЕНИЕ И КЛАССИФИКАЦИЯ *p-n* ПЕРЕХОДОВ

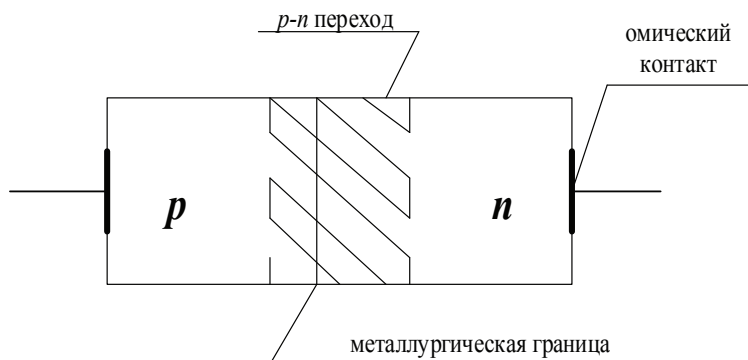
Принцип действия большинства полупроводниковых приборов основан на физических эффектах, происходящих в области контакта твердых тел. При этом преимущественно используются контакты: полупроводник—полупроводник; металл — полупроводник; металл — диэлектрик — полупроводник.

Если переход создается между полупроводниками *n*-типа и *p*-типа, то его называют электронно-дырочным, или *p-n* переходом.

Электронно-дырочный (*p-n*) переход — это обедненная подвижными носителями заряда область на границе раздела полупроводников *p*- и *n*-типа, обладающая уникальными свойствами:

- 1) содержит объемные заряды;
- 2) является потенциальным барьером;
- 3) обладает односторонней проводимостью;
- 4) становится нелинейной управляемой емкостью;
- 5) способна изменять концентрацию носителей заряда в прилегающих областях на расстоянии диффузионной длины.

В полупроводнике с областями *n*- и *p*-типов, образующими переход, можно выделить следующие пространственные области (рис. 60): металлургическую границу (воображаемая плоскость, разделяющая *p*- и *n*-области); область перехода (располагается по обе стороны металлургической границы и имеет толщину 10^{-6} — 10^{-4} см в зависимости от технологии производства); нейтральные области, лежащие между областью пространственного заряда и границами полупроводников *p*- и *n*-типов, и омические (невыпрямляющие) контакты, которыми заканчиваются нейтральные области.

Рис. 60. Структура p - n перехода

Различают переходы следующих видов:

- 1) по материалу контактирующих веществ —
 - p - n переход (обычно монокристалл);
 - p^+ - p , n^+ - n переходы — между областями с различной концентрацией примеси одного типа (верхний индекс «+» соответствует слою со значительно большей концентрацией примеси);
 - гетеропереходы — между различными полупроводниками, отличающимися шириной запрещенной зоны;
 - металл — полупроводник;
 - металл — диэлектрик — полупроводник;
- 2) по закону изменения концентрации примесей —
 - резкий или ступенчатый — переход с идеальной границей, по одну сторону которой находятся доноры с постоянной концентрацией N_d , а по другую — акцепторы с постоянной концентрацией N_A , или переход, в котором область изменения примеси много меньше ширины перехода;
 - плавный — переход, у которого в районе металлургической границы концентрация одного типа примеси постепенно уменьшается, а другого типа — растет. Сама металлургическая граница в этом случае соответствует равенству концентраций примеси ($N_A = N_d$). По закону изменения концентрации примесей плавные переходы разделяют на линейные и экспоненциальные;
- 3) по соотношению концентраций примесей в областях —
 - симметричный — $N_A \approx N_d$. Симметричные переходы не типичны для полупроводниковой техники;
 - несимметричный — $N_A \gg N_d$. (p^+ - n) или $N_A \ll N_d$. (p - n^+);

- 4) по технологии производства —
- сплавной;
 - диффузионный;
 - эпитаксиальный.

8.2. ТЕХНОЛОГИЯ ПРОИЗВОДСТВА *p-n* ПЕРЕХОДА

Получение заготовок и очистка

Заготовками для изготовления полупроводниковых приборов служат пластины германия или кремния толщиной 0,1–0,3 мм. Эти пластины должны иметь весьма малое содержание примесей (например, германий — не более 10^{17}) и обладать монокристаллической структурой. Германий и кремний обычно получают химическим путем из двуокиси. Однако полученные таким способом германий и кремний являются поликристаллическими веществами и с точки зрения требований полупроводниковой техники очень загрязнены. Поэтому особую проблему представляет очистка полупроводников и получение монокристаллов с предельными параметрами (типом проводимости, удельным сопротивлением и др.).

Методы очистки германия и кремния основаны на свойстве большинства примесей плохо переходить из жидкой фазы в твердую при кристаллизации. Если медленно охлаждать расплавленный германий или кремний, то в затвердевшей части примесей окажется меньше, чем в оставшейся жидкой. Отрезая от слитка ту часть, которая затвердела последней, можно исключить из полупроводника часть примесей. Подобные операции повторяют несколько раз и получают чистый металл. Этот принцип реализован в таких распространенных методах очистки полупроводника, как зонная плавка, выращивание кристалла из расплава (метод Чохральского) и бестигельная плавка, которые осуществляются в атмосфере инертных газов или водорода.

При зонной плавке исходный материал помещается в челночный тигель из кварца или графита, относительно которого могут перемещаться катушки индукционного нагрева (группа из 5–6 катушек). При индукционном нагреве слиток расплавляется только под катушкой, перемещением которой примеси оттесняются в жидкой фазе к одному

из краев тигеля. Одновременное перемещение нескольких катушек позволяет получить за одну «протяжку» более высокую степень очистки.

Сплавной метод

Сплавление полупроводника с металлами или их сплавами — это технологический процесс, который состоит в том, что в пластину полупроводника вплавляют металл или сплав металла, содержащий примеси, необходимые для образования зоны с электропроводностью требуемого типа. Затем систему нагревают до температуры, при которой примесь расплавится и начнется частичное растворение материала полупроводника в примесном материале. После охлаждения в полупроводнике образуется область с электропроводностью требуемого типа.

Так, при изготовлении сплавного германиевого *p-n* перехода с эмиттером *p*-типа, пластину монокристалла германия *n*-типа и таблетку индия (рис. 61, *a*) помещают в вакуумную или водородную печь и нагревают до температуры 450–550 °С, при которой таблетка с прилегающим к ней слоем пластины расплавляется, образуя сплав германий-индий. После необходимой для сплавления выдержки, нагрев прекращают. При охлаждении германий начинает переходить в твердую фазу, наращиваясь на кристаллическую решетку твердой части монокристалла. В этом тонком рекристаллизованном слое остается большое число атомов индия (10^{18} – 10^{19} см⁻³) и весь рекристаллизованный слой германия приобретает электропроводность *p*-типа. На границе рекристаллизованный слой — исходная пластина образуется резкий *p-n* переход. Рекристаллизованный слой имеет очень малое удельное сопротивление и является эмиттером по отношению к более высокоомной исходной пластине — базе.

Остальная часть застывшей капли состоит из индия, который образует с *p*-областью омический контакт. К индию припаивают вывод — обычно никелевую проволочку. Другой вывод припаивают к нанесенному на нижнюю грань исходной пластины слою олова, образующему омический контакт с германием *n*-типа (рис. 61, *б*).

Кремниевые сплавные переходы изготавливают при более высоких температурах (680–700 °С). Распространен способ сплавления алюминия в пластину кремния *n*-типа. При высокой температуре и достаточной выдержке, поверхность кремния смачивается алюминием и они взаимно растворяются (образуется сплав кремния и алюминия —

силумин). В процессе охлаждения создается рекристаллизованный слой кремния с резко выраженной дырочной проводимостью. Между этим слоем и основным кристаллом кремния n -типа образуется p - n переход. В нижнюю грань пластины вплавляют фольгу из сплава золота с сурьмой, образующую слой рекристаллизованного кремния с повышенной электронной электропроводностью (n^+ -слой). К алюминию и слою кремния n^+ -типа подсоединяют выводы. Полученный переход подвергают химическому травлению для очищения поверхности и удаления с нее закорачивающих участков, промывке и сушке. Затем наносят на него защитное покрытие.

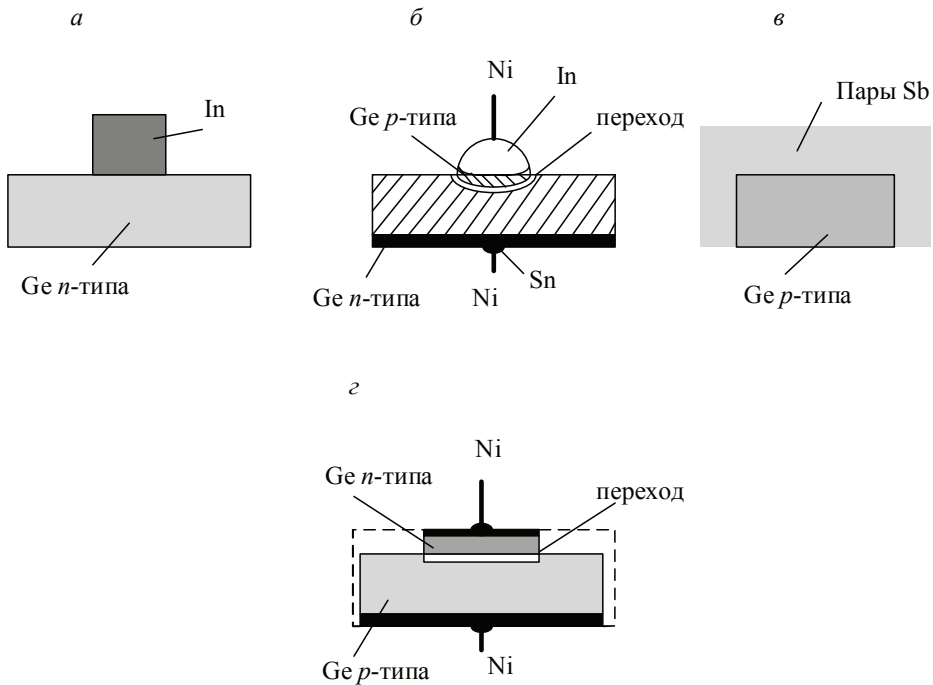


Рис. 61. Пример конструкции сплавного и диффузионного перехода

Сплавные p - n переходы относятся к числу резких (ступенчатых). Они имеют высокую надежность, работоспособность при больших обратных напряжениях, малое собственное сопротивление p - n областей, что при прямом смещении p - n перехода обеспечивает малое падение напряжения на них. Этот технологический процесс широко применяют при массовом изготовлении сплавных диодов и транзисторов.

Электрохимический метод

Электрохимические методы получения *p-n* переходов применяют, когда необходимы малые расстояния между *p*- и *n*-областями (например, в транзисторе можно получать расстояние между эмиттером и коллектором порядка 3–4 мкм). Сущность метода состоит в электрохимическом осаждении металла на поверхность полупроводника. В результате реакции образуется контакт металл — полупроводник, свойства которого зависят от физических характеристик материалов.

В редких случаях применяют комбинирование электрохимического осаждения и сплавления. Для этого полупроводник, в лунках которого произведено осаждение металла, нагревают до температуры, необходимой для сплавления последнего в полупроводник. Такую технологию создания *p-n* переходов называют микросплавной.

При производстве высокочастотных полупроводниковых приборов обычно используют *p-n* переходы, полученные электрохимическим осаждением и сплавлением.

Диффузионный метод

Для введения примесей в исходный полупроводник и получения плоскостных *p-n* переходов наиболее широко применяют высокотемпературную диффузию. Переход, образованный в результате диффузии примеси в полупроводник, называют диффузионным. При изготовлении диффузионного германиевого *p-n* перехода, в качестве исходного материала используют пластину монокристалла германия *p*-типа, а в качестве донора — сурьму (в германии коэффициент диффузии донорных примесей значительно больше, чем акцепторных). Диффузия осуществляется в водородной печи (рис. 61, *в*). Кристалл германия нагревается до температуры близкой к температуре плавления. Атомы сурьмы напыляются из газовой фазы на поверхность кристалла и диффундируют вглубь него. На поверхности образуется тонкий диффузионный слой германия с электропроводностью *n*-типа. После травления, этот слой (и соответственно *p-n* переход) оставляют только на одной из граней пластины (рис. 61, *г*). При изготовлении кремниевых переходов, в качестве исходного материала обычно берут пластину кремния с электропроводностью *n*-типа, а в качестве диффузанта — бор.

В диффузионном слое концентрация примесей уменьшается от поверхности вглубь кристалла по закону близкому к экспоненциальному. При диффузии образуется плавный *p-n* переход. Роль эмиттера выполняет высоколегированный диффузионный слой.

Переход *p-n* возникает в области, в которой концентрация носителей заряда близка к той, которая имеется у материала без примеси (при собственной электропроводности). Из-за неравномерного распределения примеси по толщине, в области, полученной диффузией, имеется собственное электрическое поле.

Разница в значениях коэффициентов диффузии у разных материалов использована для одновременного получения двух областей с разным типом электропроводности. Так, для германия коэффициент диффузии донорных примесей на несколько порядков выше коэффициента диффузии акцепторных примесей, а в кремнии наблюдается обратная картина. Поэтому, если пластину полупроводника поместить в высокотемпературную среду газа, содержащего пары как донорных, так и акцепторных примесей, атомы примесей с большим коэффициентом диффузии проникнут глубже внутрь полупроводника и создадут область с соответствующей электропроводностью. Атомы примесей с меньшим коэффициентом диффузии образуют вблизи поверхности полупроводника область с противоположным типом электропроводности. При этом необходимо, чтобы концентрация примесей с малым коэффициентом диффузии была значительно выше концентрации примеси с большим коэффициентом диффузии. Качество процесса диффузионного получения переходов во многом зависит от точности поддержания требуемой температуры. Например, при температуре 1000–1200 °С, изменение ее на несколько градусов может в два раза изменить коэффициент диффузии.

Двухстадийную (двухэтапную) диффузию применяют для уменьшения влияния изменения температуры на качество полупроводниковых приборов, получаемых методом диффузии. На первой стадии на поверхности полупроводниковой пластины при сравнительно низкой температуре получают стеклообразный слой, содержащий легирующие примеси. На второй — полупроводниковую пластину помещают в печь с более высокой температурой, при которой диффузия примесей проникает из стеклообразного слоя вглубь пластины, а на поверхности полупроводника остается диэлектрическая пленка оксида. Двухстадийный процесс диффузии часто используют при введении примесей бора

в кремний. В качестве источника примесей используется борный ангидрид B_2O_3 . Нагревая пластину и борный ангидрид в атмосфере водорода, на ее поверхности получают слой боросиликатного стекла. Нагрев пластины до более высокой температуры обеспечивает диффузию бора из слоя стекла внутрь пластины. При этом поверхность оказывается покрытой оксидом SiO_2 , который является диэлектриком. Таким образом, при двухстадийной диффузии осуществляется дозированное введение примесей из стеклообразного слоя в полупроводник.

Метод эпитаксиального наращивания

Полупроводниковая технология широко использует эпитаксиальное наращивание — создание на монокристаллической подложке слоя полупроводника, сохраняющего структуру подложки. Полупроводниковые эпитаксиальные пленки могут быть получены различными способами: термическим испарением в вакууме, осаждением из паровой фазы, распылением в газовом промежутке. Изменяя тип примеси и условия выращивания, можно в широких пределах изменять электрические свойства эпитаксиальной пленки. Наиболее распространенным способом эпитаксиального наращивания кремния является термическое восстановление его из тетрахлорида $SiCl_4$. При этом смесь тетрахлорида кремния с водородом-носителем пропускается над нагретой до $1200^\circ C$ пластиной кремния. В результате химической реакции освобождается соляная кислота HCl , пары которой удаляются из зоны реакции, а на пластину осаждается элементарный кремний, атомы которого ориентируются в соответствии с осями ее кристаллической решетки, образуя новый слой монокристалла. Если в газовую смесь ввести галоидные соединения бора или фосфора, то можно получить эпитаксиальный слой кремния не собственной, а соответственно дырочной или электронной электропроводности. С помощью эпитаксиального наращивания удобно получать однородные и тонкие (порядка нескольких микрон) слои полупроводника любого типа электропроводности и любого удельного сопротивления на более толстой пластине полупроводника тоже с любым типом электропроводности и любым удельным сопротивлением (p на p , n на n^+ , n^+ на p и другие, как это показано на рис. 62). Следует отметить, что процесс эпитаксии при изготовлении полупроводниковых элементов может заменить процесс диффузии.

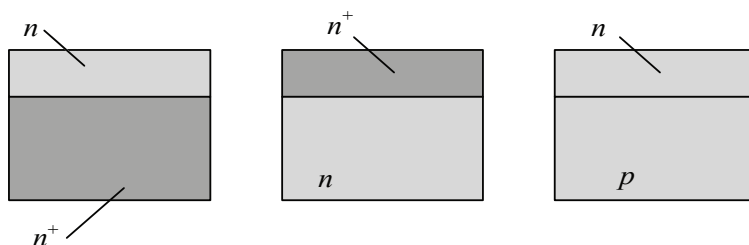


Рис. 62. Примеры эпитаксиально наращенных слоев

Переход, образованный эпитаксиальным наращиванием, называется эпитаксиальным. Из-за трудности получения резкой границы между эпитаксиальным слоем и подложкой, эпитаксиальные p - n переходы пока широко не применяются. Однако большое распространение получило эпитаксиальное наращивание тонких слоев с заданным удельным сопротивлением на толстую исходную подложку.

Метод ионного легирования

Ионное легирование сводится к бомбардировке в вакууме нагретой полупроводниковой пластины ионами примеси, ускоренными до определенной скорости. Ионы, внедрившиеся в полупроводниковую пластину, играют роль донорных или акцепторных примесей. Это позволяет, не прибегая к процессу диффузии, получать зоны, имеющие определенный тип электропроводности. Такую технологию называют ионной.

В настоящее время в производстве полупроводниковых приборов используют ионную имплантацию — легирование примесями одного из изотопов бора. При этом для маскирования используют или тонкий слой алюминия, или толстый слой диоксида кремния.

Метод вакуумного напыления

Вакуумное напыление заключается в следующем. Напыляемый металл нагревают в вакууме до температуры испарения. Затем его осаждают на покрываемую поверхность, имеющую сравнительно низкую температуру. Для получения требуемого рисунка, напыление производят через металлические маски, имеющие соответствующие прорезы.

Метод катодного распыления

Катодное распыление применяют для осаждения тугоплавких соединений. Процесс основан на явлении разрушения катода при бомбардировке его ионизированными атомами разреженного газа. Инертный газ, например аргон, вводят в испарительную камеру под давлением $1-10^2$ Па. В системе создают тлеющий разряд. Ионы газа интенсивно бомбардируют катод, в результате чего его атомы приобретают необходимую энергию и вылетают с поверхности катода. Затем они попадают на полупроводниковые пластины и, оседая на них, покрывают полупроводник слоем металла.

Метод электролитического и химического осаждения

Данный метод применяют при наличии электропроводной подложки из инертного по отношению к электролиту материала. На нее электролитическим или химическим путем осаждается пленка из водного раствора солей металлов (электролита).

Оксидное маскирование используют для того, чтобы обеспечить диффузию только в определенные участки пластины, а остальную поверхность защитить от проникновения атомов примеси. Хорошей маской, ограничивающей области диффузии, является диоксид кремния SiO_2 . Это объясняется тем, что скорость диффузии примесей в диоксиде кремния значительно меньше, чем в чистом кремнии. Кроме того, диоксид кремния является хорошим диэлектриком, поэтому окисление — неотъемлемый этап технологического процесса изготовления интегральных микросхем. Для получения оксида, пластину нагревают до температуры $900-1200$ °С в атмосфере влажного кислорода. В полученной пленке оксида согласно принципиальной схеме в последующем вытравливают окна. Этот процесс обычно применяют при изготовлении кремниевых интегральных микросхем.

Метод выращивания кристалла

По методу выращивания кристалла, в тигель с расплавленным материалом опускают «затравку» монокристалла того же химического состава и затем ее медленно поднимают вместе с нарастающим монокристаллом, одновременно слабо вибрируя или вращая ее для переме-

щения расплава. Этот метод часто применяют после зонной плавки. Методы зонной плавки и выращивания кристалла из расплава хороши для очистки германия. Для кремния, который химически очень активен вблизи температуры плавления, применяют бестигельную плавку — вариант зонной плавки, в котором исключена возможность перехода примесей из стенок тигля в кремний. При бестигельной плавке слиток подвешивают вертикально, а катушка индукционного нагрева перемещается сверху вниз. При послойном расплавлении нижняя часть слитка удерживается с верхней частью силами поверхностного натяжения. Качество получаемых монокристаллов полупроводника контролируется металлографическими испытаниями и электрическими измерениями.

После ориентации (определения в монокристалле основных кристаллических плоскостей) полученные монокристаллические слитки нарезают на множество тонких пластин либо с помощью стальных полотен, смачиваемых суспензией, в состав которой входит тонкий абразивный порошок, либо с помощью алмазных дисков, охлаждаемых струей воды. Для удаления неровностей, доводки толщины и получения параллельности поверхностей, пластины шлифуют на шлифовальных кругах. После шлифовки на поверхности остаются слои с механическими нарушениями, искажениями кристаллической решетки и остаточными механическими напряжениями. Эти слои удаляются полировкой, обычно химической, которой иногда предшествует и механическая. Химическая полировка представляет собой растворение поверхностного слоя полупроводника химическими реактивами-травителями. Перед травлением пластину очищают и обезжиривают в органических растворителях и промывают в деионизированной воде. После обработки параллельность плоскостей пластины достигает одного и даже долей микromетра на сантиметр длины. Такие пластины после тщательной очистки идут на изготовление полупроводниковых приборов.

Планарная технология

В общем случае планарным называют переход, образованный в результате диффузии примесей сквозь отверстия в защитном слое, нанесенном на поверхность полупроводника. Под планарной технологией понимают более узкую совокупность технологических операций

создания кремниевых полупроводниковых приборов и интегральных микросхем методами локальной диффузии с использованием оксидных масок — трафаретов, обеспечивающих избирательную защиту отдельных участков пластины — подложки. Оксидную маску получают методом фотолитографии. Некоторые вещества, находящиеся на поверхности полупроводника, препятствуют диффузии в него ряда примесей. Слой двуокиси SiO_2 толщиной 0,5–1,0 мкм препятствует диффузии бора и фосфора. Слой (пленку) двуокиси получают термическим окислением кремния либо в атмосфере сухого или влажного кислорода, либо в парах воды при температуре выше 1000 °С. Далее методом фотолитографии в отдельных местах пленки образуют отверстия нужных конфигураций и размеров.

Фотолитография — это процесс получения на поверхности пластины требуемого рисунка. Сущность фотолитографии заключается в следующем (рис. 63).

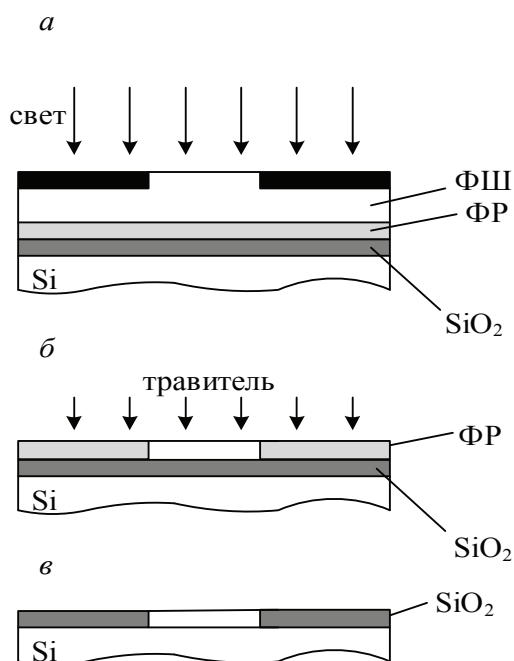


Рис. 63. Основные этапы фотолитографии

На слой двуокиси наносят слой фотоэмульсии, так называемого фоторезиста. После сушки слой фоторезист ФР засвечивают ультра-

фиолетовым светом через стеклянный или кварцевый фотошаблон ФШ с белыми и черными тонами, соответствующими рисунку оксидной маски (рис. 63, *а*). После удаления в проявителе незасвеченных участков фоторезиста (или засвеченных в зависимости от того, применяется негативный или позитивный фоторезист), оставшийся фоторезист закрепляют (дублирование). Далее протравливают слой двуокиси кремния (вплоть до поверхности полупроводника) в тех местах, где он не защищен слоем фоторезиста (рис. 63, *б*). После травления остатки фоторезиста удаляют химическим путем (рис. 63, *в*). Таким образом на поверхности пластины кремния получают оксидную маску с необходимой совокупностью отверстий — окон, через которые в дальнейшем и производится диффузия примесей.

При изготовлении самого перехода (рис. 64) исходной заготовкой служит пластина кремния n^+ -типа 1 с очень высокой удельной электропроводностью (для получения малого объемного сопротивления базы). На одной из граней пластины путем эпитаксиального наращивания получают тонкий (толщиной в несколько микрон) слой кремния 2 того же типа электропроводности, но с большим удельным сопротивлением для получения высокого пробивного напряжения перехода. Нанеся на него оксидную маску 3, переходят к высокотемпературной диффузии акцепторов (обычно бора или алюминия). Проникая сквозь окно к поверхности полупроводника, примесные атомы образуют под этим окном область кремния p -типа 4 и p - n переход 5. Во время диффузии окно, через которое осуществлялась диффузия, вновь покрывается тонкой пленкой SiO_2 . Новая пленка защищает поверхность перехода от влияния различных внешних факторов (влаги, загрязнения). В этой пленке двуокиси методом фотолитографии вновь делают окно, через которое на поверхность p -области осаждается алюминий 6. Алюминий далее сплавляется с кремнием (получается омический контакт с p -областью). К алюминиевой пленке методом термокомпрессии (сильным прижатием к пленке алюминия тонкой алюминиевой проволочки при температуре 200–300 °С) подсоединяют один из выводов. Другой вывод подсоединяют к омическому контакту, образованному с основной пластиной кремния.

Особенностью планарной технологии является то, что рабочие области полупроводникового прибора сложной структуры выходят на одну поверхность. Поэтому все выводы могут располагаться на одной грани

кристалла, что удобно при конструктивном оформлении транзисторов и интегральных микросхем.

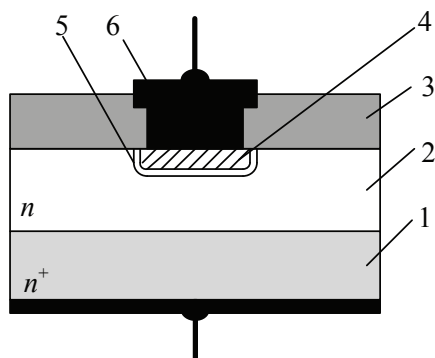


Рис. 64. Структура планарного перехода

Точечные переходы

Точечным называют переход, все размеры которого меньше толщины области пространственного заряда. Эти переходы появились раньше плоскостных.

Точечный переход образуется в месте контакта тонкого (диаметром 10–20 мкм) острия металлической иглы с кристаллом германия (редко кремния) в результате формовки. Через контакт пропускают один или несколько сравнительно мощных, но коротких импульсов тока. При этом приконтактная область сильно разогревается, кончик иглы сплавляется с полупроводником и контакт приобретает стабильность и механическую прочность. Процесс формовки сопровождается изменением типа электропроводности примыкающего к игле тонкого слоя исходного полупроводника. Образуется полусферический *p-n* переход (рис. 65).

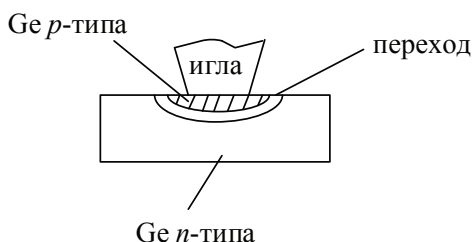


Рис. 65. Структура точечного перехода

Изменение типа электропроводности происходит в основном за счет диффузии примесей из иглы в полупроводник во время формовки, когда игла и пластина полупроводника у контакта частично расплавляются. Например, при контакте иглы из бериллиевой бронзы с германием n -типа, в германий диффундирует акцептор — бериллий. Для изготовления точечных переходов обычно применяют полупроводники n -типа, т. к. точечный контакт плохо формируется к полупроводнику p -типа.

8.3. РАВНОВЕСНОЕ СОСТОЯНИЕ p - n ПЕРЕХОДА

Образование p - n перехода

Рассмотрим образование несимметричного p - n перехода при идеальном контакте двух полупроводников с различным типом проводимости (рис. 66). Через плоскость металлургического контакта (плоскость, где изменяется тип примесей, преобладающих в полупроводниках) возникает диффузия из-за градиента концентрации носителей заряда. В результате диффузии носителей заряда нарушается электрическая нейтральность примыкающих к металлургическому контакту частей монокристалла полупроводника.

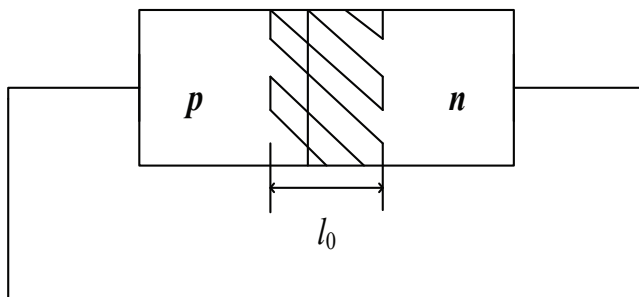


Рис. 66. Равновесный несимметричный p - n переход

На рис. 67 указано, что внешнее напряжение на переход не подается, а p - и n -области соединены между собой внешним проводником, устанавливающим p - n переход в равновесное состояние.

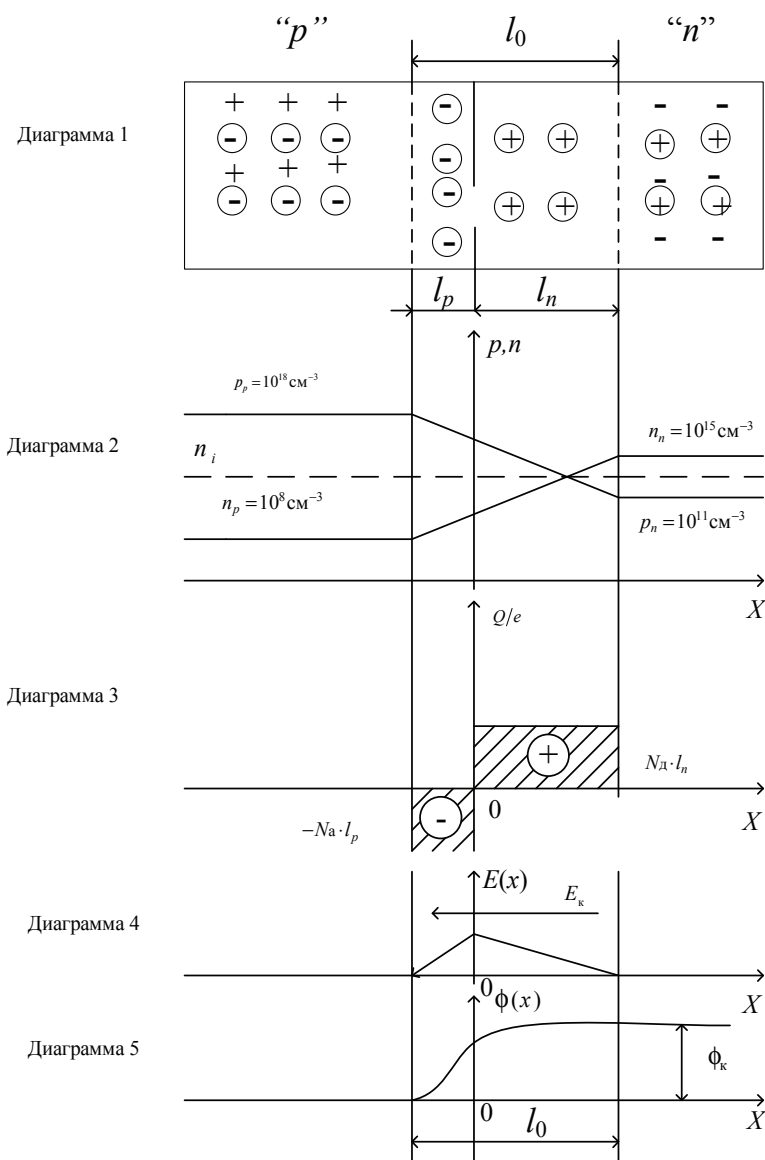


Рис. 67. Образование несимметричного p-n перехода:

+ — дырка — основной носитель заряда полупроводника p-типа; — — электрон — основной носитель заряда полупроводника n-типа; \oplus — положительный ион донора; \ominus — отрицательный ион акцептора; l_p — ширина p-n перехода в области полупроводника p-типа; l_n — ширина p-n перехода в области полупроводника n-типа; l_0 — ширина p-n перехода в равновесном состоянии

Пусть концентрация акцепторов N_A в области полупроводника p -типа больше концентрации доноров N_D в области полупроводника n -типа $N_A \gg N_D$. При этом концентрация основных носителей заряда — дырок — в полупроводнике p -типа будет больше концентрации основных носителей заряда — электронов — в полупроводнике n -типа $p_p > n_n$. Соответственно концентрация неосновных носителей заряда — электронов — в полупроводнике p -типа меньше концентрации неосновных носителей заряда — дырок — в полупроводнике n -типа $n_p < p_n$.

Допустим, $N_A = 10^{18} \text{ см}^{-3}$, а $N_D = 10^{15} \text{ см}^{-3}$. Поясним процесс образования p - n перехода с помощью диаграмм, представленных на рис. 67.

Распределение концентраций основных и неосновных носителей заряда в полупроводниках определяется из закона действующих масс. Так, для полупроводника p -типа, закон действующих масс записывается в виде

$$n_i^2 = p_p \cdot n_p = N_A \cdot n_p.$$

Допустим, что для изготовления p - n перехода используется полупроводниковый материал германий, у которого собственная концентрация (концентрация свободных носителей заряда в полупроводнике i -типа) носителей заряда составляет величину $n_{iGE} = 2,5 \cdot 10^{13} \text{ см}^{-3}$. При условии $p_p = N_A = 10^{18} \text{ см}^{-3}$ из закона действующих масс находим, что

$$n_p = n_i^2 / N_A = 6,25 \cdot 10^{26} / 10^{18} = 6,25 \cdot 10^8 \text{ см}^{-3}.$$

В полупроводнике n -типа закон действующих масс определяется по соотношению

$$n_i^2 = n_n \cdot p_n = N_D \cdot p_n.$$

При условии $n_n = N_D = 10^{15} \text{ см}^{-3}$ из закона действующих масс получаем, что $p_n = 6,25 \cdot 10^{11} \text{ см}^{-3}$.

В результате разности концентраций подвижных носителей заряда, на границе контакта полупроводников p - и n -типов (диаграмма 2 рис. 67) существует градиент концентрации носителей заряда каждого знака. Под действием градиента концентрации будет происходить диффузия основных носителей заряда из области с высокой концентрацией в область с меньшей их концентрацией. Поскольку концентрация дырок в p -области выше, чем в n -области, то часть дырок в результате диффузии перейдет в n -область, вблизи границы которой окажутся избыточные дырки, которые будут рекомбинировать с элект-

тронами. В результате в этой зоне уменьшится концентрация свободных электронов и образуется область некомпенсированных положительных ионов донорной примеси.

В p -области уход дырок из граничного слоя способствует образованию области с некомпенсированными отрицательными зарядами ионов акцепторной примеси. Подобным же образом происходит диффузионное перемещение электронов из n -слоя в p -слой (диаграмма 1 на рис. 67). Аналогично этому и электроны из области полупроводника n -типа переходят в область полупроводника p -типа, оставляя в полупроводнике n -типа положительные ионы доноров. В области полупроводника p -типа, при рекомбинации электронов с дырками, дополнительно открываются отрицательные ионы акцепторов.

Отрицательные ионы акцепторов и положительные ионы доноров находятся в узлах кристаллической решетки и поэтому не могут двигаться по кристаллу полупроводника. Область образовавшихся неподвижных пространственных зарядов (ионов) и есть область p - n перехода (диаграмма 3 на рис. 67). В ней имеют место пониженная концентрация основных носителей заряда и, следовательно, повышенное сопротивление, которое определяет электрическое сопротивление всей системы. В зонах, прилегающих к месту контакта двух разнородных областей, нарушается условие электронейтральности. Но за пределами p - n перехода все заряды взаимно компенсируют друг друга, и полупроводник остается электрически нейтральным.

Таким образом, вблизи контакта полупроводников с различным типом проводимости возникает двойной слой пространственного заряда: отрицательный в области полупроводника p -типа; положительный в области полупроводника n -типа (диаграмма 1 на рис. 67).

В области объемных зарядов мала концентрация подвижных носителей заряда, поэтому слой обладает повышенным сопротивлением. Он называется запиорным слоем, или p - n переходом.

Итак, электронно-дырочный, или p - n переход — это тонкий слой полупроводника, возникающий на границе раздела двух полупроводников с разным типом проводимости, который обеднен подвижными носителями заряда и обладает высоким сопротивлением.

Электрическое поле, возникающее между разноименными ионами, препятствует перемещению основных носителей заряда (диаграмма 4 на рис. 67). Поэтому поток дырок из p - в n -область и электронов

из n - в p -область уменьшается с ростом напряженности электрического поля. Однако это поле не препятствует движению через переход неосновных носителей, имеющихся в p - и n -областях. Такие носители заряда собственной электропроводности, имеющие энергию теплового происхождения, генерируются в объеме полупроводника и, диффундируя к электрическому переходу, захватываются электрическим полем. Они перебрасываются в область с противоположной электропроводностью.

Перенос неосновных носителей приводит к уменьшению объемного заряда и электрического поля в переходе. Как следствие, имеет место дополнительный диффузионный переход основных носителей, в результате чего электрическое поле принимает исходное значение. При равенстве потоков основных и неосновных носителей заряда и соответственно токов наступает динамическое равновесие.

Ширина p - n перехода может быть найдена при интегрировании уравнения Пуассона, которое определяет распределение напряженности электрического поля $E(x)$ и потенциала $\phi(x)$. При этом получают

$$l_0 = l_p + l_n = \sqrt{\frac{2\varepsilon\varepsilon_0\phi_k}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)},$$

где ε — диэлектрическая проницаемость полупроводника; ε_0 — диэлектрическая проницаемость вакуума (диэлектрическая постоянная); ϕ_k — контактная разность потенциалов; q — заряд электрона; N_A — концентрация акцепторов; N_D — концентрация доноров.

Поскольку $N_A \gg N_D$, постольку $l_p \ll l_n$, и приближенно можно записать

$$l_0 \approx l_n \approx \sqrt{\frac{2\varepsilon\varepsilon_0\phi_k}{q} \frac{1}{N_D}}.$$

Поскольку полупроводник в целом нейтрален, объемные заряды в обеих частях p - n перехода равны (диаграмма 3 на рис. 67 отражает плоское сечение объемных зарядов), т.е. $Q_p = qN_A l_p S$, $Q_n = qN_D l_n S$, где S — площадь p - n перехода. Учитывая, что $Q_n = Q_p$ или $N_A \cdot l_p = N_D \cdot l_n$, получим соотношение $N_A/N_D = l_n/l_p$, которое означает, что во сколько раз концентрация акцепторов больше концентрации доноров, во столько же раз ширина p - n перехода в полупроводнике n -типа больше ширины p - n перехода в полупроводнике p -типа. В нашем случае $N_A \gg N_D$, а $l_n \gg l_p$ и p - n переход в основном находится в области полупроводни-

ка *n*-типа, т.е. в высокоомном слое. Такие *p-n* переходы называются несимметричными переходами. Обычно ширина *p-n* перехода l_0 равна 0,1–1,0 мкм.

Распределение напряженности электрического поля и потенциала в *p-n* переходе (диаграммы 4 и 5 на рис. 68) получают из решения уравнения Пуассона

$$\frac{dE(x)}{dx} = -\frac{d^2\varphi(x)}{dx^2} = -\frac{\lambda(x)}{\varepsilon_0 \cdot \varepsilon},$$

где $\lambda(x)$ — плотность объемного заряда.

Распределение плотности объемного заряда $\lambda(x)$ в *p-n* переходе аппроксимируют функцией $\lambda(x) = -qN_A$ при $x \in [0, l_p]$. Интегрируя данное уравнение Пуассона с учетом того, что на границах *p-n* перехода $E = 0$, получают распределение напряженности электрического поля $E(x)$, которое выражается кусочно-линейной функцией (диаграмма 4 на рис. 67), поскольку ее производная $\lambda(x)$ постоянна на участках $(-l_p, 0)$ и $(0, l_n)$. Максимальная напряженность наблюдается на металлургической границе контакта двух полупроводников

$$E_{\text{макс}} \approx \sqrt{\frac{2qN_d\varphi_K}{\varepsilon_0\varepsilon}}.$$

При увеличении концентрации примеси возрастает максимальное значение напряженности электрического поля в *p-n* переходе. Электрическое поле препятствует переходу основных носителей заряда через *p-n* переход. При контакте двух полупроводников возникает потенциальный барьер (распределение потенциала вдоль *p-n* перехода показано на диаграмме 5 рис. 67). Функция $\varphi(x)$ также получается путем двойного интегрирования уравнения Пуассона, причем она состоит из двух параболических участков, поскольку получена в результате интегрирования кусочно-линейной функции $E(x)$ и имеет точку перегиба при $x = 0$.

Высота потенциального барьера в равновесном состоянии равна контактной разности потенциалов φ_K

$$\varphi_K = \varphi_p - \varphi_n,$$

где φ_p — потенциал работы выхода электрона дырочного полупроводника; φ_n — потенциал работы выхода электрона электронного полупроводника.

Для определения высоты потенциального барьера, интегрирование $E(x)$ проводится с учетом граничных условий: $\varphi_p = 0$ при $x = l_p$; $\varphi_n = -\varphi_k$ при $x = l_n$.

Поскольку значение φ_k отрицательно и потенциальный барьер имеет отрицательное значение, то можно сделать вывод, что потенциальный барьер препятствует диффузии основных носителей заряда. Электрическое поле, возникающее в p - n переходе в процессе его образования, способствует переходу неосновных носителей заряда: электронов в полупроводнике p -типа; дырок в полупроводнике n -типа.

8.4. Ток в p - n ПЕРЕХОДЕ В РАВНОВЕСНОМ СОСТОЯНИИ

Ток в полупроводниках определяется направленным движением носителей заряда. Направленное движение носителей заряда имеет две разновидности: диффузионное направленное движение носителей заряда и дрейфовое направленное движение носителей заряда. Первое направленное движение носителей заряда — диффузия — вызывается неравномерной концентрацией носителей заряда в полупроводнике, а второе — дрейф — происходит под действием электрического поля. В p - n переходе присутствуют обе составляющие направленного движения носителей заряда, и ток в p - n переходе содержит две составляющие: диффузионную составляющую тока p - n перехода, дрейфовую составляющую тока p - n перехода.

Ток диффузии обусловлен переходом основных носителей заряда через p - n переход из-за их градиента концентрации (за счет тепловой энергии). Переход основных носителей заряда происходит в тормозящем электрическом поле p - n перехода. Плотность диффузионного тока определяется по закону Фика

$$i_D = i_{Dp} + i_{Dn} = -(qD_p \frac{dp}{dx} + qD_n \frac{dn}{dx}),$$

где i_D — плотность диффузионного тока p - n перехода; i_{Dp} — дырочная составляющая плотности тока диффузии; i_{Dn} — электронная составляющая плотности тока диффузии; D_p , D_n — коэффициенты диффузии дырок и электронов; dp/dx , dn/dx — градиенты концентрации дырок и электронов. Знак минус перед скобкой показывает, что диффузион-

ный ток протекает в сторону уменьшения градиента концентрации.

Учтем, что рассматриваемый p - n переход является несимметричным, т. е. $N_A \gg N_D$, а значит, и $p_p \gg n_n$; из этого следует, что $i_{Dp} \gg i_{Dn}$.

Дрейфовая составляющая тока обусловлена переносом через p - n переход неосновных носителей заряда под действием ускоряющего для них электрического поля p - n перехода напряженностью E_k . Плотность дрейфового тока i_E определяется

$$i_E = \sigma E$$

или

$$i_E = i_{Ep} + i_{En} = q\mu_p p_n E + q\mu_n n_p E,$$

где i_{Ep} — дырочная составляющая плотности дрейфового тока; i_{En} — электронная составляющая плотности дрейфового тока; μ_p, μ_n — подвижность дырок и электронов; p_n — концентрация дырок в полупроводнике n -типа, являющихся там неосновными носителями заряда; n_p — концентрация электронов в полупроводнике p -типа, являющихся там также неосновными носителями заряда.

Исходя из первоначального предположения, что $N_A \gg N_D$, и закона действующих масс, можно сделать вывод, что $p_n \gg n_p$, поэтому $i_{Ep} \gg i_{En}$.

Непременным условием равновесного состояния p - n перехода является равенство нулю суммарного тока через p - n переход. Направления токов в p - n переходе в равновесном состоянии показаны на рис. 68.

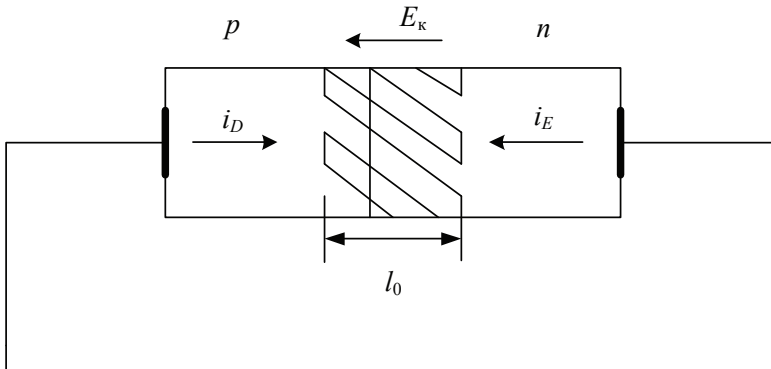


Рис. 68. Токи в равновесном p - n переходе

Ток в p - n переходе в основном определяется дырочной составляющей тока диффузии полупроводника p -типа и дырочной составляющей дрейфового тока полупроводника n -типа. Поэтому для плотности

тока через p - n переход следует записать условие равновесного состояния p - n перехода в виде $i_D + i_E = 0$, иначе $i_{Dn} + i_{Dp} + i_{En} + i_{Ep} = 0$, или приближенно $i_{Dp} + i_{Ep} = 0$, т. к. $i_{Dp} \gg i_{Dn}$, $i_{Ep} \gg i_{En}$.

8.5. КОНТАКТНАЯ РАЗНОСТЬ ПОТЕНЦИАЛОВ

Распределение потенциала в p - n переходе находится из решения уравнения Пуассона. Максимальное значение потенциала, определяющего высоту потенциального барьера для равновесного состояния, равно контактной разности потенциалов (диаграмма 5 на рис. 67). Ее можно определить на основании следующих рассуждений.

Условие равновесия p - n перехода имеет вид $i_D + i_E = 0$.

Поскольку и диффузионная, и дрейфовая составляющие плотности тока определяются в основном соответствующими дырочными составляющими i_{Dp} и i_{Ep} , то условие равновесия можно приближенно записать $i_{Dp} + i_{Ep} = 0$, или в развернутом виде

$$-qD_p \frac{dp_p}{dx} + q\mu_p p_n E = 0.$$

Из последнего выражения с учетом $E = dU/dx$ запишем

$$D_p \cdot \frac{dp_p}{dx} = p_n \cdot \mu_p \cdot E = p_n \cdot \mu_p \frac{dU}{dx}.$$

Преобразуя данную формулу, получим $dU = \frac{D_p \cdot dp_p}{\mu_p \cdot p_n}$,

или в общем виде $dU = \left(\frac{D}{\mu}\right) \left(\frac{dp}{p}\right)$.

Параметры дрейфового и диффузионного движения носителей заряда связаны между собой соотношениями Эйнштейна:

$$D_n = \frac{kT}{q} \cdot \mu_n = \Phi_T \cdot \mu_n; \quad D_p = \frac{kT}{q} \cdot \mu_p = \Phi_T \cdot \mu_p.$$

В общем случае $D = \frac{kT}{q} \cdot \mu = \Phi_T \cdot \mu$, где D — коэффициент диффузии;

μ — подвижность. Отсюда находим

$$D/\mu = kT/q = \varphi_T,$$

где φ_T — температурный потенциал. Для комнатной температуры $T = 300 \text{ K}$ имеем $\varphi_T = kT/q \approx 25 \text{ мВ}$. Подставляя φ_T в соотношения Эйнштейна, получим дифференциальное уравнение $dU = \varphi_T \cdot dp/p$. Решение данного дифференциального уравнения в результате интегрирования в общем виде представляется в виде соотношения

$$U = \varphi_k = \varphi_T \ln(p) + C.$$

Для определения постоянной C необходимо знание граничных условий. Для полупроводника p -типа граничные условия запишем в виде $p = p_p$; имеем φ — потенциал работы выхода электрона из полупроводника p -типа, $\varphi = \varphi_p$, $\varphi_p = \varphi_T \ln(p_p)$. Для полупроводника n -типа граничные условия имеют вид $p = p_n$, φ — потенциал работы выхода электрона из полупроводника n -типа, $\varphi = \varphi_n$, $\varphi_n = \varphi_T \ln(p_n)$. Отсюда контактная разность потенциалов запишется как

$$\varphi_k = \varphi_p - \varphi_n = \varphi_T \ln(p_p) - \varphi_T \ln(p_n) = \varphi_T \ln\left(\frac{p_p}{p_n}\right).$$

С учетом закона действующих масс, для полупроводника n -типа имеем $p_n = n_i^2/n_n \approx n_i^2/N_d$, т. к. $n_n = n_i + N_d \approx N_d$, $N_d \gg n_i$.

В результате подстановки этих соотношений в выражение для контактной разности потенциалов получим

$$\varphi_k = \varphi_T \ln(n_n \cdot p_p / n_i^2) = \frac{kT}{q} \ln\left(\frac{N_d \cdot p_p}{n_i^2}\right) = \frac{kT}{q} \ln\left(\frac{N_d \cdot N_A}{n_i^2}\right),$$

т. к. $p_p = n_i + N_A \approx N_A$, $N_A \gg n_i$ ($n_i = p_i$).

Окончательная формула для контактной разности потенциалов p - n перехода определяется ее зависимостью от трех факторов:

$$\varphi_k = f[\text{материал полупроводника}; N_{\text{пр}}; t, ^\circ\text{C}].$$

Зависимость φ_k от материала полупроводника определяется различным значением их ширины запрещенной зоны. Известно, что для германия $\Delta W_z = 0,72 \text{ эВ}$, $n_i \approx 10^{13} \text{ см}^{-3}$; для кремния $\Delta W_z = 1,12 \text{ эВ}$, $n_i \approx 10^{10} \text{ см}^{-3}$; для арсенида галлия $\Delta W_z = 1,41 \text{ эВ}$, $n_i \approx 10^6 \text{ см}^{-3}$. При внесении одинаковой концентрации примеси во все полупроводниковые материалы и изготовлении из них p - n переходов, на основании формулы контактной разности потенциалов следует, что $\varphi_{\text{кGe}} < \varphi_{\text{кSi}} < \varphi_{\text{кGaAs}}$, т. е. чем больше ширина запрещенной зоны полупроводника, тем больше контактная разность потенциалов. При комнатной температуре ори-

ентировочные значения контактной разности потенциалов p - n переходов из различных полупроводниковых материалов составляют

$$\varphi_{\text{кGe}} = 0,3...0,4 \text{ В}, \varphi_{\text{кSi}} = 0,6...0,8 \text{ В}, \varphi_{\text{кGaAs}} = 1,0...1,2 \text{ В}.$$

Степень легирования исходных полупроводников также влияет на значение контактной разности потенциалов. Чем больше степень легирования полупроводника, т. е. чем больше вносится в полупроводник атомов примеси ($N_{\text{пр}}$ — концентрация примеси), тем большее значение имеет контактная разность потенциалов.

Контактная разность потенциалов зависит от температуры окружающей среды. С увеличением температуры контактная разность потенциалов уменьшается. Это связано с тем, что в выражении для $\varphi_{\text{к}}$ с увеличением температуры окружающей среды возрастает значение температурного потенциала φ_T , но $n_i = AT^{3/2} \exp(-\frac{\Delta W_3}{2kT})$ также возрастает, и это увеличение происходит быстрее, чем рост температурного потенциала, поэтому контактная разность потенциалов при увеличении температуры уменьшается.

8.6. ЭНЕРГЕТИЧЕСКАЯ ДИАГРАММА p - n ПЕРЕХОДА В РАВНОВЕСНОМ СОСТОЯНИИ

В условиях равновесия p - n перехода, когда отсутствует внешнее напряжение, энергия Ферми одинакова для любого объема полупроводника, поэтому уровень Ферми занимает горизонтальное положение на энергетической диаграмме (рис. 69).

Уровень Ферми в полупроводнике p -типа расположен вблизи энергетического уровня потолка валентной зоны, а в полупроводнике n -типа — вблизи энергетического уровня дна зоны проводимости, причем уровень Ферми ближе расположен к энергетическому уровню потолка валентной зоны, чем к энергетическому уровню дна зоны проводимости из-за того, что $N_A \gg N_D$. У изолированных p - и n -областей энергии Ферми не равны, поэтому, при объединении областей в единый кристалл полупроводника, на основании фундаментального свойства уровня Ферми ($\text{grad } W_F = 0$) происходит смещение энергетических уровней n -области относительно энергетических уровней p -области, как и показано на рис. 69.

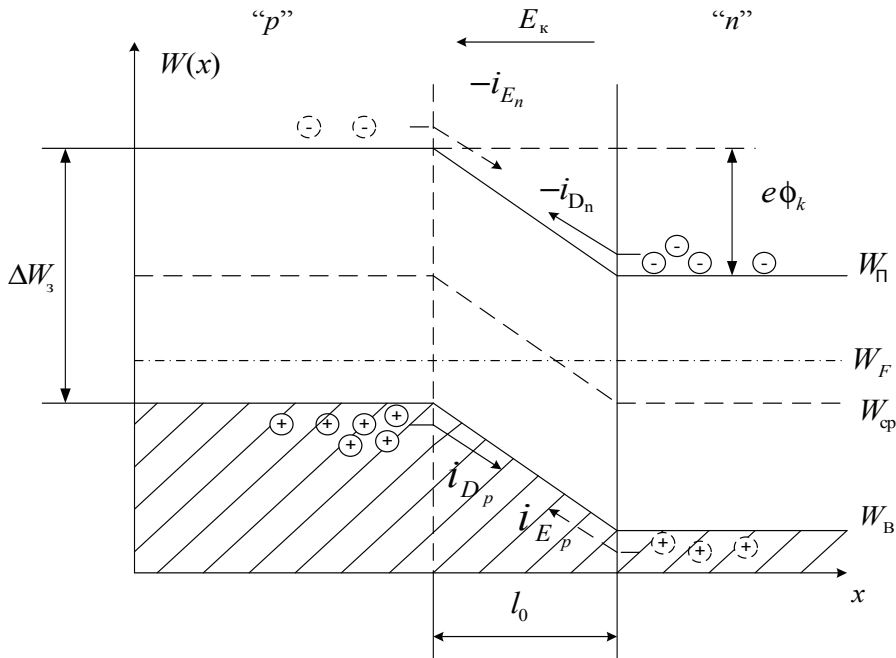


Рис. 69. Энергетическая диаграмма равновесного p - n перехода:

⊖ ⊕ — основные носители заряда; ⊖ ⊕ — неосновные носители заряда;
 W_n — энергетический уровень дна зоны проводимости; W_F — энергетический
 уровень Ферми; W_{cp} — энергетический уровень середины запрещенной зоны;
 W_B — энергетический уровень потолка валентной зоны; ΔW_3 — энергия,
 соответствующая ширине запрещенной зоны

В результате смещения энергетических уровней создается энергетический (потенциальный) барьер

$$q\phi_k = W_{Fn} - W_{Fp},$$

где W_{Fn} — энергия Ферми для полупроводника n -типа; W_{Fp} — энергия Ферми для полупроводника p -типа.

Основные носители заряда областей полупроводника p - и n -типов, энергия которых больше высоты барьера, диффузионно преодолевают его. Основные носители заряда, переходящие p - n переход в тормозящем для них электрическом поле, образуют диффузионную составляющую тока перехода i_D . В то же время неосновные носители заряда, находящиеся вблизи p - n перехода и совершающие тепловое хаотическое движение, попадают под действие электрического поля p - n перехода, которое переносит их в противоположную область: электро-

ны p -области в n -область; дырки n -области в p -область. Неосновные носители заряда, преодолевающие переход под действием напряженности электрического поля E_k p - n перехода, образуют дрейфовую составляющую тока i_E через переход. Условие равновесия выполняется, когда диффузионный ток i_D будет скомпенсирован встречным дрейфовым током i_E и полный ток через переход будет равен нулю $i_D + i_E = 0$.

Легко увидеть по энергетической диаграмме p - n перехода влияние температуры на величину потенциального барьера (рис. 70).

При увеличении температуры окружающей среды, на основании свойства уровня Ферми, его положение изменяется и становится ближе к энергии середины запрещенной зоны W_{cp} как в полупроводнике p -типа, так и в полупроводнике n -типа. Это смещение уровня Ферми, при увеличении температуры в p - и n -областях, происходит в противоположных направлениях (рис. 70), что неизбежно привело бы к появлению $\text{grad} W_F > 0$, а это недопустимо, т. к., на основании свойства уровня Ферми, в условии равновесия уровень Ферми должен находиться в горизонтальном положении, т. е. на одном энергетическом уровне как в p -области, так и в n -области.

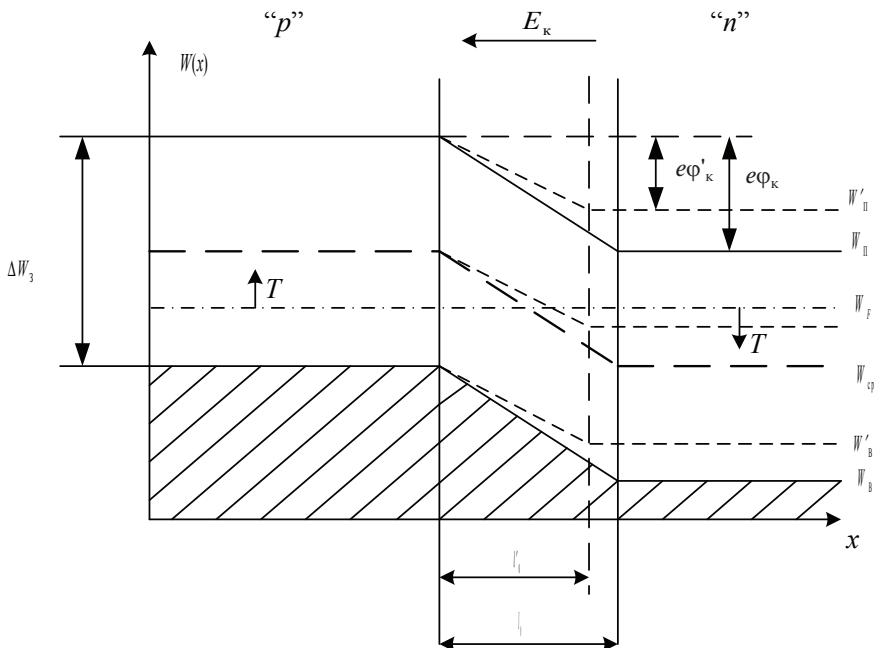


Рис. 70. Изменение энергетической диаграммы при повышении температуры

В то же время, при увеличении температуры, ширина запрещенной зоны почти не изменяется (температурный коэффициент ширины запрещенной зоны составляет около -10^{-4} 1/К), поэтому с ростом температуры происходит смещение энергетических уровней границ разрешенных зон *n*-области и уменьшается высота энергетического барьера, что ведет к снижению потенциального барьера *p-n* перехода, а равно и к уменьшению контактной разности потенциалов.

На рис. 70 сплошной линией изображено положение энергетических уровней при $T_1 = 300$ К (контактная разность потенциалов φ_k , равновесная ширина *p-n* перехода l_0), пунктирной линией показано изменение положений энергетических уровней при $T_2 = 340$ К (новая контактная разность потенциалов $\varphi'_k < \varphi_k$, а ширина *p-n* перехода $l'_0 < l_0$). На диаграмме показана система в равновесном состоянии *p-n* перехода и при соблюдении условия $\text{grad } W_F = 0$, $N_A \gg N_D$.

8.7. НЕРАВНОВЕСНОЕ СОСТОЯНИЕ *p-n* ПЕРЕХОДА

Если подключить источник ЭДС U между *p*- и *n*-слоями, то равновесие перехода нарушится и в цепи потечет ток. Поскольку удельное сопротивление обедненного слоя намного выше, чем удельное сопротивление нейтральных слоев, то внешнее напряжение практически полностью падает на переходе и изменяет высоту потенциального барьера.

Прямосмещенный p-n переход

Если к *p*-области подсоединить положительный полюс внешнего источника напряжения, а к *n*-области — отрицательный, такое включение *p-n* перехода будет прямым смещением *p-n* перехода. В этом случае под действием внешнего электрического поля основные носители заряда начнут перемещаться в сторону *p-n* перехода.

При подаче внешнего напряжения на *p-n* переход изменяется его ширина, что видно из потенциальной диаграммы рис. 71. Ширину прямосмещенного *p-n* перехода можно найти из выражения

$$l_{\text{пр}} = l_0 \cdot \sqrt{\frac{\varphi_k - U_{\text{пр}}}{\varphi_k}},$$

где l_0 — ширина p - n перехода в равновесном состоянии,

$$l_0 = \sqrt{\frac{2\varepsilon_0\varepsilon\phi_k}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)}.$$

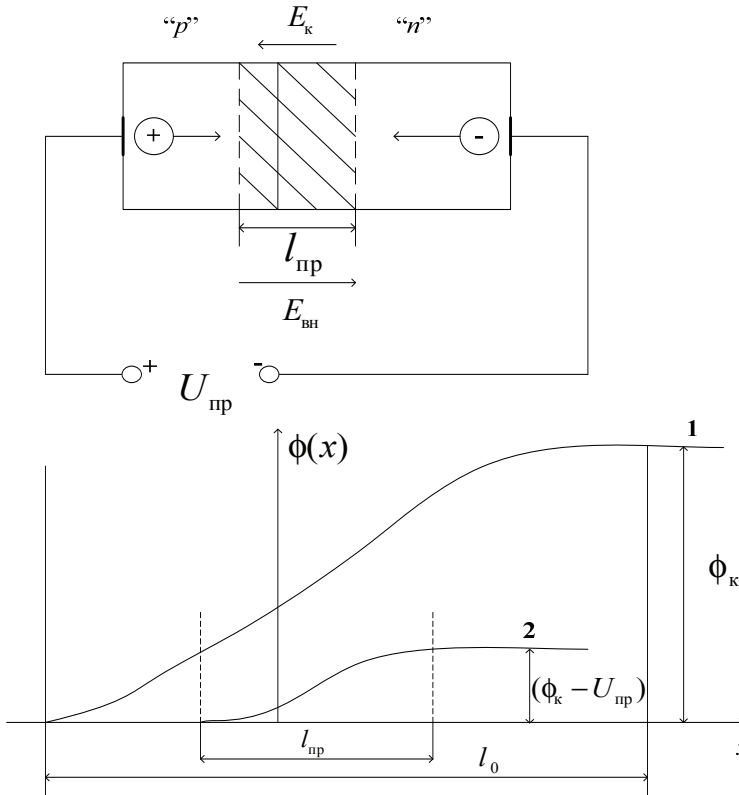


Рис. 71. Схема включения и диаграмма распределения потенциала для прямосмещенного p - n перехода:

$U_{пр}$ — напряжение внешнего источника, приложенного к p - n переходу в прямом направлении (прямое смещение p - n перехода); $E_{вн}$ — напряженность внешнего электрического поля; 1 — распределение потенциала вдоль p - n перехода в равновесном состоянии; 2 — распределение потенциала вдоль p - n перехода при прямом смещении; l_0 — ширина p - n перехода в равновесном состоянии; $l_{пр}$ — ширина p - n перехода при подаче прямого напряжения $U_{пр}$

Из потенциальной диаграммы рис. 71 следует, что при прямом смещении высота потенциального барьера снижается

$$\phi_{пр} = \phi_k - U_{пр}.$$

В результате частичной компенсации внешним напряжением потенциального барьера происходит резкое увеличение тока диффузии через переход $i_D = i_{Dp} + i_{Dn}$, т. к. все больше основных носителей заряда оказываются способными преодолеть уменьшившийся потенциальный барьер. Следовательно, существовавший в равновесном состоянии баланс токов диффузии и дрейфа нарушается и через переход потечет ток, который называется прямым.

В несимметричном *p-n* переходе, ток диффузии создается в основном потоком дырок из *p*-области в *n*-область, т. к. встречный поток электронов мал и им можно пренебречь ($N_A = 10^{18} \text{ см}^{-3} \gg (N_D = 10^{15} \text{ см}^{-3})$; $i_{Dp} \gg i_{Dn}$). При этом в *n*-области существенно возрастает концентрация избыточных неосновных носителей заряда — дырок, перешедших из *p*-области. Данное образование избыточной концентрации носителей заряда получило название инжекции.

Инжекция — процесс нагнетания (введения) носителей заряда через *p-n* переход при понижении высоты потенциального барьера в область полупроводника, для которого они являются неосновными носителями заряда. В несимметричном переходе инжекция имеет односторонний характер. Область, инжектирующая носители заряда, называется эмиттером. Эта область сильно легирована примесями и имеет низкое удельное электрическое сопротивление. Область, в которую инжектируются неосновные для нее носители заряда, называется базой. База меньше легирована примесями и имеет большее значение удельного электрического сопротивления.

В области базы инжектированные из эмиттера дырки за счет наличия градиента концентрации распространяются вглубь базы, рекомбинируют с электронами базы, при этом концентрация дырок уменьшается, приближаясь к равновесной.

Закон изменения концентрации дырок в базе описывается функцией

$$p_n = p_{n0} + \Delta p \exp\left(-\frac{x}{L_p}\right),$$

где p_n — концентрация дырок в базе; p_{n0} — равновесная концентрация дырок в базе; Δp — избыточная концентрация дырок на границе базы с *p-n* переходом; L_p — диффузионная длина дырок в базе.

Избыточная концентрация дырок в базе является функцией приложенного прямого напряжения $\Delta p = f(U_{np})$. На рис. 72 точка *A* соответствует диффузионной длине дырок в базе.

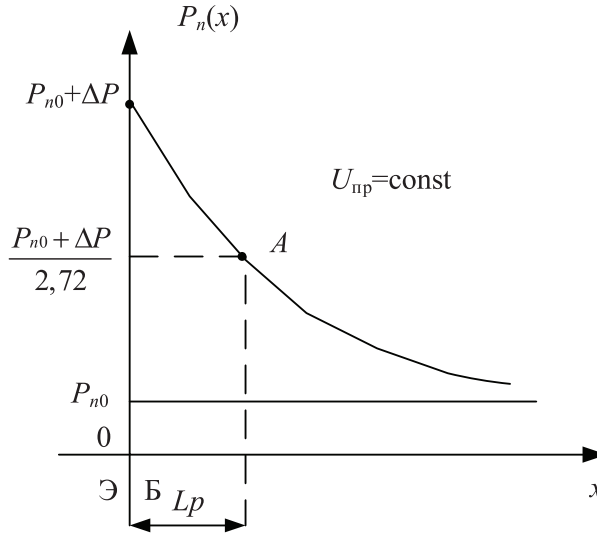


Рис. 72. Избыточная концентрации дырок в базе

Отрезок, на котором избыточная концентрация неосновных носителей заряда (дырок) уменьшается в $e(2,72)$ раз вследствие рекомбинации, называется длиной диффузии L_p . На участке длиной $(3...4)L$ избыточная концентрация уменьшается 20–50 раз, т. е. становится пренебрежимо малой по сравнению с исходной. Учитывая выражение для диффузионной скорости

$$V_{\text{диф}} = D/L = L/\tau,$$

можно дать еще одно определение L : диффузионная длина есть то расстояние, которое носители заряда в процессе диффузии проходят за время жизни (τ).

В результате инжекции на границах перехода в p - и n -областях окажутся дополнительные носители заряда, неосновные для данной области. Вблизи p - n перехода концентрации дырок в n -области и электронов в p -области оказывается больше равновесной концентрации, определяемой по закону действующих масс

$$p_n = p_{n_0} \left[\exp(U_{\text{пр}}/\phi_T) \right], \quad n_p = n_{p_0} \left[\exp(U_{\text{пр}}/\phi_T) \right].$$

В таком случае в каждом слое появляются избыточные носители (т. е. происходит инжекция)

$$\Delta p_n = p_{n_0} \left[\exp(U_{\text{пр}}/\phi_T) - 1 \right], \quad \Delta n_p = n_{p_0} \left[\exp(U_{\text{пр}}/\phi_T) - 1 \right].$$

Следовательно, избыточные концентрации неосновных носителей заряда на границе *p-n* перехода увеличиваются по экспоненциальному закону в зависимости от напряжения, приложенного к нему.

Дополнительные неосновные носители заряда в течение времени $(3...5)\tau_e$ компенсируются основными носителями заряда, которые приходят из объема полупроводника (где τ_e — время диэлектрической релаксации). В результате этого на границе *p-n* перехода появляется заряд, созданный основными носителями заряда, и выполняется условие $\Delta n_n \approx \Delta p_n$, $\Delta p_p \approx \Delta n_p$. Электронеutrальность полупроводника восстанавливается. Такое перераспределение основных носителей заряда приводит к появлению электрического тока во внешней цепи, т. к. по ней поступают носители заряда взамен ушедших к *p-n* переходу и исчезнувших в результате рекомбинации.

Приведенные выше соотношения справедливы для условия низкого уровня инжекции. Уровень инжекции δ — отношение концентрации избыточных носителей (инжектированных) к равновесной концентрации основных носителей базы, т. е. $\delta = \frac{\Delta p_n}{n_{n0}} \approx \frac{\Delta p_n}{N_D}$. Низкому уровню ин-

жекции соответствует неравенство $\delta \ll 1$, т. е. $\Delta p_n \approx \Delta n_n \ll n_{n0}$.

При прямом смещении (рис. 73) уровень Ферми полупроводника в *n*-области смещается вверх относительно его положения в *p*-области на величину qU_{np} . На эту же величину снижается высота энергетического барьера. При этом дрейфовая составляющая тока *p-n* перехода не изменяется, т. к. условия перехода неосновных носителей заряда через *p-n* переход остаются теми же, что и в равновесном состоянии, т. е. переход неосновных носителей заряда происходит в ускоряющем электрическом поле *p-n* перехода. Из-за снижения высоты энергетического барьера, количество переходов основных носителей заряда в тормозящем электрическом поле *p-n* перехода будет резко увеличиваться, а следовательно, возрастет диффузионная составляющая тока перехода.

Дрейфовая составляющая тока, при приложении прямого напряжения, остается практически без изменения. Это обусловлено тем, что создающие такой ток неосновные носители генерируются вблизи *p-n* перехода на расстоянии, меньшем диффузионной длины L (где L — среднее расстояние, на которое носитель диффундирует за время жизни). Те заряды, которые рождаются на большом расстоянии,

в основном рекомбинируют, не дойдя до перехода. Изменение ширины перехода для носителей заряда этого происхождения не играет существенной роли. Они как генерировались в пределах толщины, определяемой диффузионной длиной, так и будут генерироваться. Следовательно, ток, обусловленный движением указанных носителей заряда, останется без изменения — таким же, как и в равновесном состоянии.

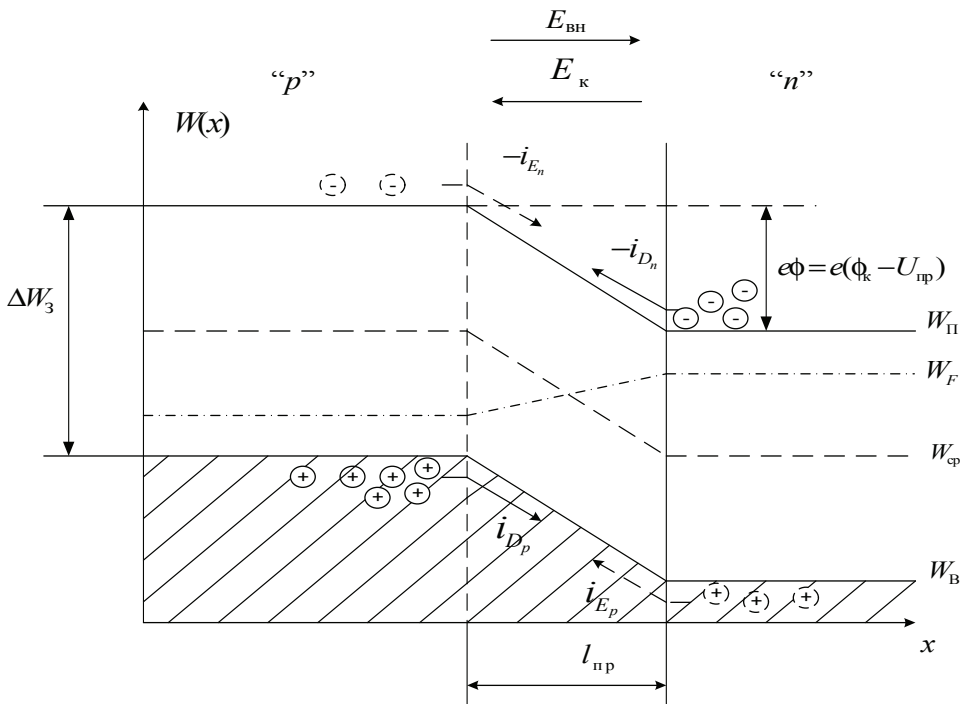


Рис. 73. Энергетическая диаграмма прямосмещенного p - n перехода

Таким образом, для несимметричного ($N_a \gg N_d$) прямосмещенного p - n перехода характерны следующие соотношения между токами:

$$i_{Dp} \gg i_{Dn}, i_D \gg i_E.$$

Обратносмещенный p - n переход

Если к p -области подключить отрицательный полюс внешнего источника напряжения, а к n -области — положительный, то та-

кое включение p - n перехода будет обратным смещением p - n перехода (рис. 74).

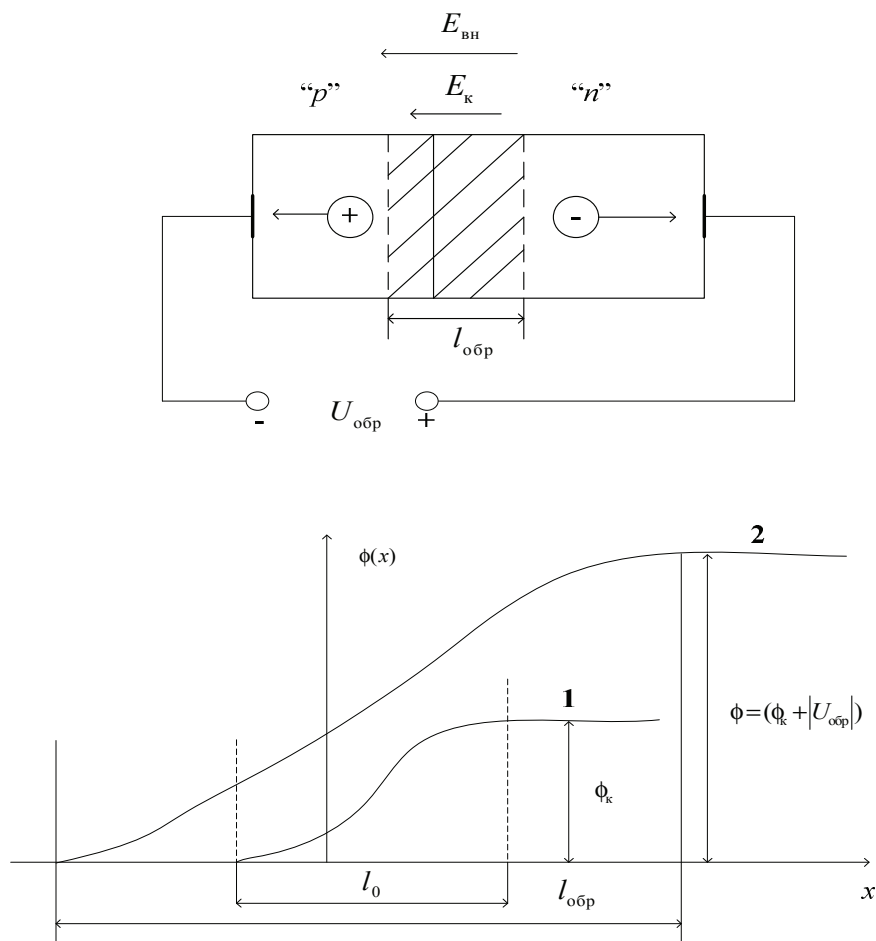


Рис. 74. Схема включения и диаграмма распределения потенциала для обратносмещенного p - n перехода:

l_0 — ширина p - n перехода в равновесном состоянии; 1 — распределение потенциалов в равновесном состоянии p - n перехода; $l_{обр}$ — ширина p - n перехода при обратном смещении; 2 — потенциальная диаграмма при обратном смещении p - n перехода; ϕ_k — высота потенциального барьера в равновесном состоянии p - n перехода; $(\phi_k + |U_{обр}|)$ — высота потенциального барьера при обратном смещении p - n перехода

Под действием обратного напряжения $U_{обр}$, основные носители заряда будут перемещаться от границ p - n перехода вглубь обла-

стей. При этом ширина p - n перехода увеличивается. Она определяется по формуле

$$l_{\text{обр}} = l_0 \cdot \sqrt{\frac{\varphi_{\text{к}} + |U_{\text{обр}}|}{\varphi_{\text{к}}}} \approx l_0 \cdot \sqrt{\frac{U_{\text{обр}}}{\varphi_{\text{к}}}}.$$

Приближенная запись для $l_{\text{обр}}$ оправдана, т. к. $|U_{\text{обр}}| \gg \varphi_{\text{к}}$. Из формулы для $l_{\text{обр}}$ видно, что p - n переход расширяется нелинейно с увеличением приложенного напряжения $U_{\text{обр}}$: сначала быстрее, затем расширение p - n перехода замедляется.

При подаче $U_{\text{обр}}$ увеличивается потенциальный барьер, т. к. напряженность внешнего электрического поля $E_{\text{вн}}$ совпадает с направлением напряженности внутреннего электрического поля $E_{\text{к}}$, уменьшается число основных носителей заряда, способных его преодолеть, и ток диффузии уменьшается. Уже при $|U_{\text{обр}}| = (0,1 \dots 0,2) \varphi_{\text{к}}$ ток диффузии становится равным нулю, а через p - n переход протекает лишь ток неосновных носителей заряда, образующих дрейфовую составляющую тока.

В результате действия обратного напряжения, снижается концентрация неосновных носителей заряда у границ p - n перехода и появляется их градиент концентрации. Возникает диффузия неосновных носителей заряда к границам p - n перехода, где они подхватываются электрическим полем p - n перехода и переносятся через p - n переход. Это поясняется диаграммой распределения концентраций основных и неосновных носителей заряда в областях p - n перехода.

Как показано на рис. 75, концентрация неосновных носителей заряда на границах p - n перехода практически падает до нуля. Снижение концентраций неосновных носителей заряда у границ p - n перехода, появление градиента их концентрации и диффузия неосновных носителей заряда к p - n переходу характеризуются экстракцией.

Экстракцией называется извлечение неосновных носителей заряда из областей, прилегающих к p - n переходу на расстоянии диффузионной длины, которое возникает при увеличении высоты потенциального барьера.

Энергетическая диаграмма обратносмещенного p - n перехода приведена на рис. 76.

При обратном смещении, уровень Ферми полупроводника в n -области смещается вниз относительно его положения в p -области на величину $qU_{\text{обр}}$. На эту же величину соответственно увеличивается высота энергетического барьера.

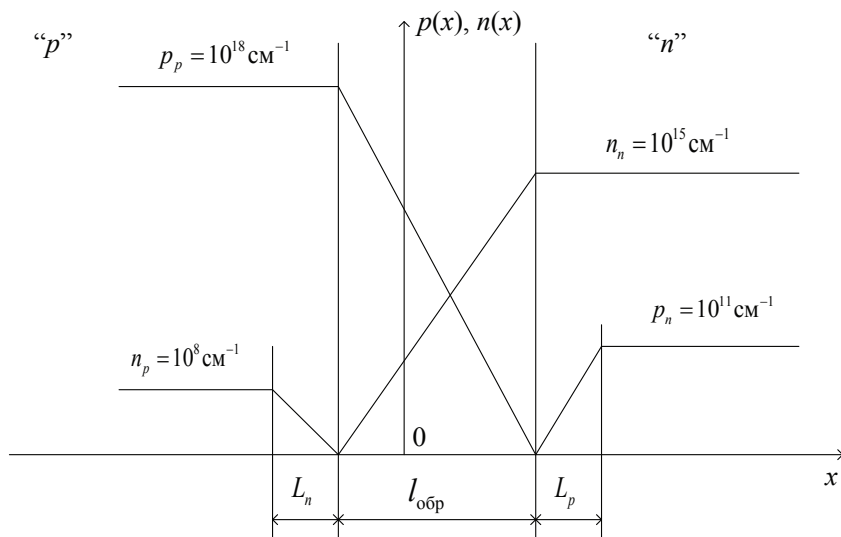


Рис. 75. Распределение концентраций носителей заряда в обратносмещенном p - n переходе (L_n, L_p — длина диффузии электронов и дырок)

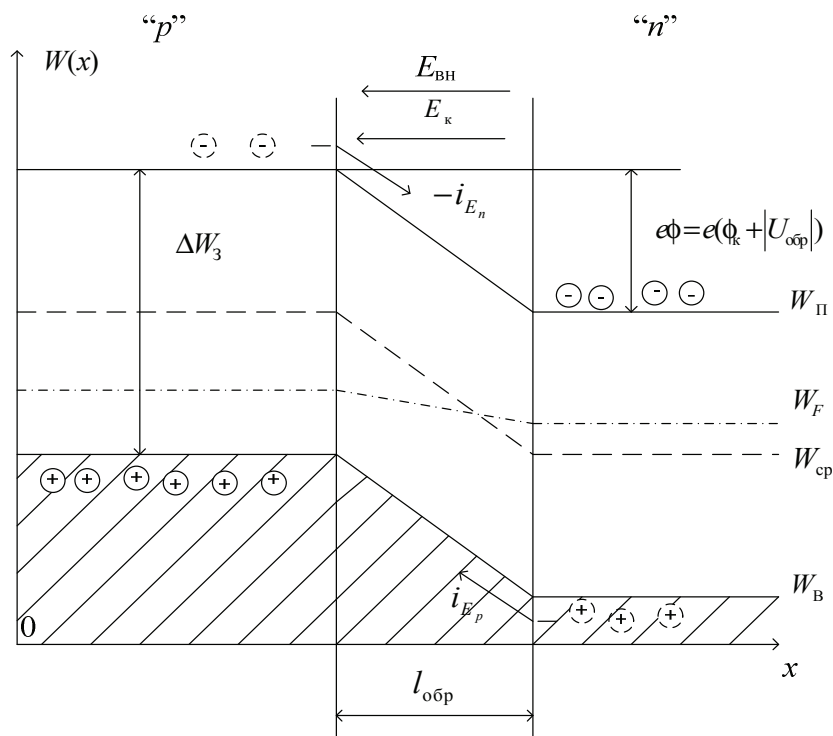


Рис. 76. Энергетическая диаграмма обратносмещенного p - n перехода

Обратный ток включает дрейфовую составляющую

$$i_E = i_{Ep} + i_{En} = qp_n \cdot \mu_p E + qn_p \cdot \mu_n E.$$

С учетом принятых значений концентраций основных и неосновных носителей имеем $p_n \gg n_p$ и $i_{Ep} \gg i_{En}$, а следовательно, можно приближенно записать $i_E \gg i_{Ep}$. Таким образом, дрейфовый ток несимметричного p - n перехода создается по большей части неосновными носителями заряда базы (слаболегированная область p - n перехода), т. е. преобладает экстракция неосновных носителей заряда из базы.

8.8. ИДЕАЛЬНЫЙ p - n ПЕРЕХОД

Основные соотношения для идеального p - n перехода

Идеальным, или идеализированным, p - n переходом называется такой p - n переход, который представляет собой упрощенную модель реального p - n перехода, для которой приняты следующие допущения:

- 1) все подводимое напряжение приложено и падает непосредственно на p - n переходе (пренебрегаем падением напряжения на прилегающих к p - n переходу областях);
- 2) ширина p - n перехода пренебрежимо мала (в p - n переходе отсутствуют явления генерации и рекомбинации электронно-дырочных пар);
- 3) характеристики p - n перехода рассматриваются при обратных напряжениях, которые значительно меньше пробивных значений;
- 4) границы p - n перехода являются плоскими, носители заряда движутся только в направлении, перпендикулярном этим границам, краевые эффекты не учитываются;
- 5) толщина нейтральных областей много больше диффузионной длины неосновных носителей заряда в этих областях;
- 6) уровень инжекции мал, т. е. при любом значении тока через переход концентрация основных носителей заряда много больше концентрации неосновных носителей заряда.

В модели p - n перехода также предполагается, что изменение концентрации неосновных носителей заряда в областях за границами перехода при небольшом прямом напряжении не нарушает электрической нейтральности данных областей. Это объясняется быстрой (за время

диэлектрической релаксации) нейтрализацией заряда инжектированных неосновных носителей заряда основными носителями заряда, поступающими из внешней цепи.

Физические процессы при прямом напряжении на p - n переходе поясним с помощью рис. 77, на котором показаны направления движения основных носителей заряда, создающих прямой ток.

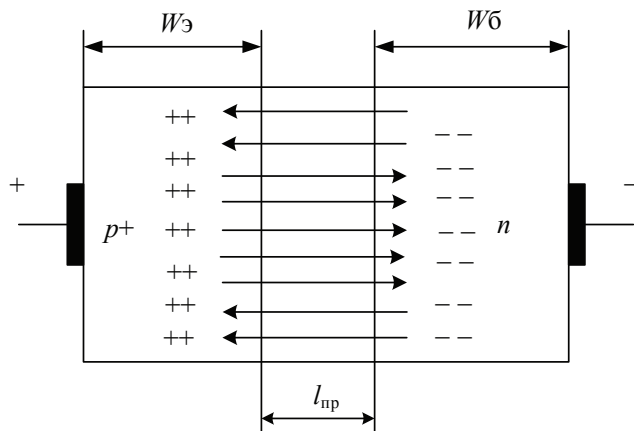


Рис. 77. Схема движения основных носителей заряда:

$W_{\text{э}}$ — толщина области эмиттера; $W_{\text{б}}$ — толщина области базы (обычно $W_{\text{б}} > W_{\text{э}}$);
 p^+ — большее превышении концентрации акцепторных примесей
 в эмиттерной области над концентрацией донорных примесей в базовой области
 (рассматривается несимметричный p - n переход)

Перемещение основных носителей заряда через p - n переход (см. рис. 77) приводит к инжекции избыточных неосновных носителей заряда — электронов — в нейтральную p -область, а дырок — в n -область. В нейтральных областях около p - n перехода, неосновные носители заряда движутся от его границ вглубь областей вследствие диффузии, вызванной возникшим градиентом концентрации неосновных носителей заряда. Стационарные распределения избыточной концентрации неосновных носителей заряда в этих областях определяются из уравнения диффузии вида

$$D_n \cdot \frac{d^2 n_p}{dx^2} - \frac{(n_p - n_{p0})}{\tau_n} = 0,$$

где D_n — коэффициент диффузии электронов; n_p — неравновесная концентрация электронов в полупроводнике p -типа; n_{p0} — равновес-

ная концентрация электронов в полупроводнике p -типа; τ_n — время жизни электрона.

Это дифференциальное уравнение — уравнение диффузии — описывает диффузионное движение электронов в дырочном полупроводнике с учетом рекомбинации.

Решение уравнения диффузии проводится при условиях, что на границах p - n перехода, принятых за начало отсчета, концентрации неосновных носителей заряда определяются следующим образом. Высота энергетического (потенциального) барьера p - n перехода в равновесном состоянии находится из соотношения

$$q\varphi_K = kT \ln \left(\frac{n_{n0}}{n_{p0}} \right) = kT \ln \left(\frac{p_{p0}}{p_{n0}} \right),$$

а при прямом смещении в этом выражении производится замена φ_K на $(\varphi_K - U_{np})$ и n_{p0} на $n_p = n_{p0} + \Delta n_p$, где Δn_p — избыточная концентрация электронов — неосновных носителей заряда в области полупроводника p -типа. В таком случае запишем

$$q(\varphi_K - U_{np}) = kT \ln \left(\frac{n_{n0}}{n_{p0} + \Delta n_p} \right) = kT \ln \left(\frac{n_{n0}}{n_{p0}} \right) - kT \ln \left(1 + \frac{\Delta n_p}{n_{p0}} \right).$$

Отсюда граничное условие для p -области будет

$$\Delta n_p = n_{p0} \left[\exp \left(\frac{U_{np}}{\varphi_T} \right) - 1 \right].$$

Аналогично для n -области граничное условие запишется как

$$\Delta p_n = p_{n0} \left[\exp \left(\frac{U_{np}}{\varphi_T} \right) - 1 \right],$$

где Δp_n — избыточная концентрация дырок — неосновных носителей заряда в области полупроводника n -типа.

Решение уравнения диффузии для p -области имеет вид

$$\Delta n_p(x) = \Delta n_p(0) \cdot \exp(-x/L_n),$$

а для n -области —

$$\Delta p_n(x) = \Delta p_n(0) \cdot \exp(-x/L_p),$$

где L_n , L_p — диффузионные длины электронов в p -области и дырок в n -области соответственно.

Распределение концентраций $\Delta n_p(x) = n_p(x) - n_{p0}$; $\Delta p_n(x) = p_n(x) - p_{n0}$ приведено на рис. 78.

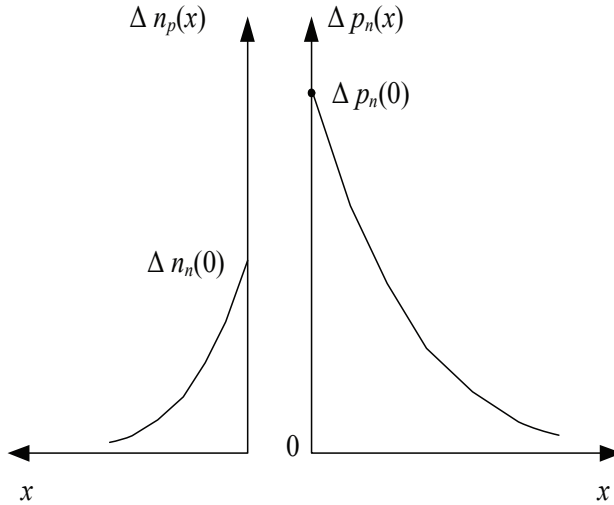


Рис. 78. Распределение избыточных концентраций носителей заряда

На рис. 78 приняты разные начала отсчета и направления отсчета координаты x : от соответствующих границ p - n перехода.

Вольт-амперная характеристика идеального p-n перехода

Под вольт-амперной характеристикой (ВАХ) p - n перехода понимается зависимость тока p - n перехода от значения приложенного к нему напряжения $I = f(U)$.

Плотность тока диффузии на границе p - n перехода определяется по соотношениям

$$i_{Dn}(0) = qD_n \cdot \left. \frac{dn_p}{dx} \right|_{x=0} = qD_n \cdot \frac{\Delta n_p(0)}{L_n};$$

$$i_{Dp}(0) = qD_p \cdot \left. \frac{dp_n}{dx} \right|_{x=0} = qD_p \cdot \frac{\Delta p_n(0)}{L_p}.$$

Ток через p - n переход состоит из электронной и дырочной составляющих токов

$$I = I_n + I_p,$$

где I_n — электронная составляющая тока p - n перехода; I_p — дырочная составляющая тока p - n перехода. На границе p - n перехода можно определить

$$I_n = S \cdot i_{Dn}(0) = qSD_n \cdot \Delta n_p(0) / L_n ,$$

$$I_p = S \cdot i_{Dp}(0) = qSD_p \cdot \Delta p_n(0) / L_p ,$$

где S — площадь p - n перехода; D_n, D_p — коэффициенты диффузии электронов в p -области и дырок в n -области.

Если подставить в выражение для I_n и I_p избыточные концентрации электронов $\Delta n_p(0)$ на границе p -области и дырок $\Delta p_n(0)$ на границе n -области исходя из соответствующих граничных условий для модели идеального p - n перехода, то в целом ток через p - n переход опишется вольт-амперной характеристикой идеального перехода в виде

$$I = I_n + I_p = I_0 (\exp[U/\phi_T] - 1) ,$$

где I_0 — тепловой ток, который оценивается из соотношения

$$\begin{aligned} I_0 &= qSL_p \cdot p_{n0} / \tau_p + qSL_n \cdot n_{p0} / \tau_n \approx \\ &\approx qSn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right). \end{aligned}$$

Здесь учтено, что $n_{n0} \approx N_D$ и $p_{p0} \approx N_A$ а τ_n, τ_p — время жизни электронов в p -области и дырок в n -области; U — напряжение на p - n переходе, оно задается со знаком «плюс» при прямом смещении и со знаком «минус» при обратном включении p - n перехода; ϕ_T — температурный потенциал, $\phi_T = kT/q$.

Тепловой ток I_0 представляет собой дрейфовую составляющую тока p - n перехода, которая образуется при движении неосновных носителей заряда через p - n переход в его ускоряющем электрическом поле. Для несимметричных p - n переходов, I_0 представляет собой ток экстракции носителей из базы. Произведение теплового тока на экспоненциальный множитель представляет диффузионную составляющую тока через p - n переход — ток инжекции носителей в базу.

Параметром, характеризующим относительную роль главной составляющей тока в переходе, является коэффициент инжекции γ . В случае несимметричного перехода (p^+ - n) данный коэффициент записывается следующим образом:

$$\gamma = \frac{I_p(0)}{I(0)} = \frac{I_p(0)}{I_p(0) + I_n(0)} \approx 1 - \frac{\rho_{\text{Э}}}{\rho_{\text{Б}}}.$$

Коэффициент инжекции показывает, какая часть полного тока создается носителями эмиттера. Для несимметричного переходов $\gamma \approx 0,95 \dots 0,999$.

Теоретическая ВАХ p-n перехода представлена на рис. 79.

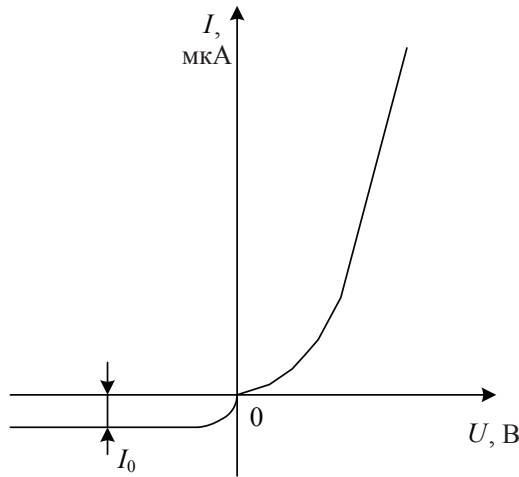


Рис. 79. Теоретическая ВАХ p-n перехода

При малых прямых напряжениях, прямой ток имеет значение близкое к значению теплового тока, но уже при $U_{\text{пр}} \approx 2,3\phi_T$ (приблизительно 60 мВ при $T = 300 \text{ K}$) $I > 10I_0$.

При подаче обратного напряжения, диффузионная составляющая тока резко уменьшается и $I_0 \exp [U_{\text{обр}}/\phi_T] \approx 0$ при $|U_{\text{обр}}| = (2 \dots 3)\phi_T$, а величина обратного тока определяется значением теплового тока $I = -I_0$, далее обратный ток не зависит от величины $U_{\text{обр}}$. Поэтому ток I_0 называют током насыщения. Если $N_{\text{д}} = 10^{16} \text{ см}^{-3}$ и $L_p = 10 \text{ мкм}$, то для кремния при комнатной температуре плотность тока насыщения $i_0 \approx 2 \cdot 10^{-10} \text{ А/см}^2$. У современных интегральных транзисторов площади баз составляют не более $2 \cdot 10^{-5} \text{ см}^2$, а площади эмиттеров — до 10^{-6} см^2 и менее. Поэтому при комнатной температуре типичным значением тока насыщения у интегрального кремниевого перехода можно считать $I_0 \approx 10^{-15} \text{ А}$. Небольшие изменения температуры приводят к резкому росту тока насыщения, поэтому его называют

тепловым током и упрощенно определяют исходя из следующих соотношений

$$I_0 = B S \exp(-\Delta W_3 / kT),$$

$$I_0 = \frac{I_{\text{пр}}}{\exp(U_{\text{пр}} / \phi_T)},$$

где B — коэффициент, зависящий от материала полупроводника; S — площадь p - n перехода.

Положение ВАХ идеального p - n перехода зависит от температуры окружающей среды и степени легирования областей p - n перехода (рис. 80).

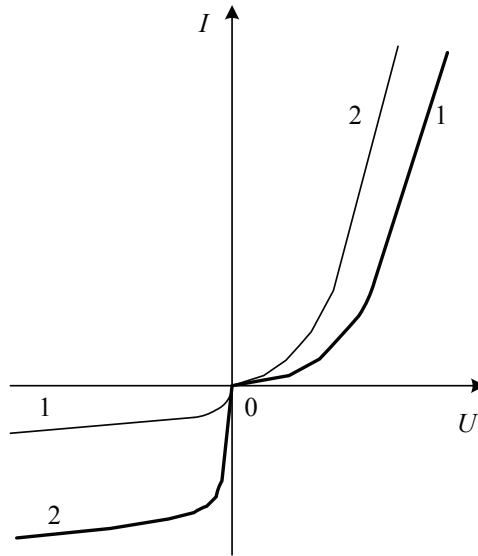


Рис. 80. Влияние температуры на ВАХ идеального p - n перехода:
1 — $T_1 = +20\text{ }^{\circ}\text{C}$; 2 — $T_2 = +40\text{ }^{\circ}\text{C}$

Ток I_0 с увеличением температуры растет экспоненциально, т. к. экспоненциально возрастает концентрация неосновных носителей заряда p_n в области полупроводника n -типа, поскольку p_n в основном и определяет величину I_0 для несимметричного перехода $p_n = n_i^2 / n_n \approx \approx \text{const} \exp[-\Delta W_3 / (kT)]$. На практике влияние температуры на ток I_0 оценивают приближенно: ток насыщения увеличивается в два раза при возрастании температуры на каждые $10\text{ }^{\circ}\text{C}$.

Прямая ветвь ВАХ p - n перехода с ростом температуры смещается влево. Это обусловлено тем, что при увеличении температуры, во-первых, возрастает тепловая энергия основных носителей заряда и растет число носителей заряда, энергия которых больше высоты энергетического (потенциального) барьера p - n перехода, во-вторых, снижается высота потенциального барьера и большее число основных носителей заряда создают прямой ток. Рост прямого тока, при увеличении температуры, в уравнении ВАХ определяется в основном возрастанием тока I_0 .

Влияние концентрации примесей областей p - n перехода на ВАХ можно пояснить на основании закона действующих масс для каждой области p - n перехода. Так, для полупроводника p -типа имеем

$$n_p = n_i^2 / p_p \approx n_i^2 / N_A ,$$

а для полупроводника n -типа —

$$p_n = n_i^2 / n_n \approx n_i^2 / N_D .$$

Известно, что ток I_0 зависит от концентрации примесей

$$I_0 = q S n_i^2 \cdot \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) ,$$

откуда следует, что с ростом N_A и N_D (степени легирования) происходит уменьшение тока I_0 .

Нелинейность ВАХ удобно оценивать, сопоставляя его сопротивления в прямом и обратном смещении. Как и для других нелинейных элементов, различают дифференциальные сопротивления и сопротивления постоянному току.

Один из важнейших параметров прямой ветви ВАХ — дифференциальное сопротивление перехода. Значение этого сопротивления легко получить из уравнения идеальной ВАХ

$$r_{\text{пер}} = dU/dI = \varphi_T / I .$$

Физический смысл дифференциального сопротивления становится ясным, если заменить дифференциалы dU и dI конечными приращениями, тогда

$$\Delta U = \Delta I r_{\text{пер}} .$$

Значит, $r_{\text{пер}}$ есть сопротивление для приращений тока ΔI малых по сравнению с постоянной составляющей тока I , определяющей ве-

личину $r_{\text{пер}}$, другими словами, $r_{\text{пер}}$ — сопротивление перехода переменному току в заданной точке ВАХ.

Дифференциальное сопротивление прямосмещенного перехода обратно пропорционально току p - n перехода.

Сопротивление перехода постоянному току $R_{\text{пер}}$ в соответствии с законом Ома определяется из ВАХ перехода

$$R_{\text{пер}} = \frac{U}{I} = \frac{\varphi_T}{I} \ln\left(\frac{I}{I_0} + 1\right).$$

На обратной ветви ВАХ, когда $|U| \gg \varphi_T$, сопротивление перехода постоянному току прямо пропорционально напряжению $R_{\text{пер}} = |U_{\text{обр}}|/I_0$. Дифференциальное сопротивление может считаться бесконечно большим $r_{\text{пер}} \rightarrow \infty$.

Графический способ определения сопротивлений перехода иллюстрируется на рис. 81.

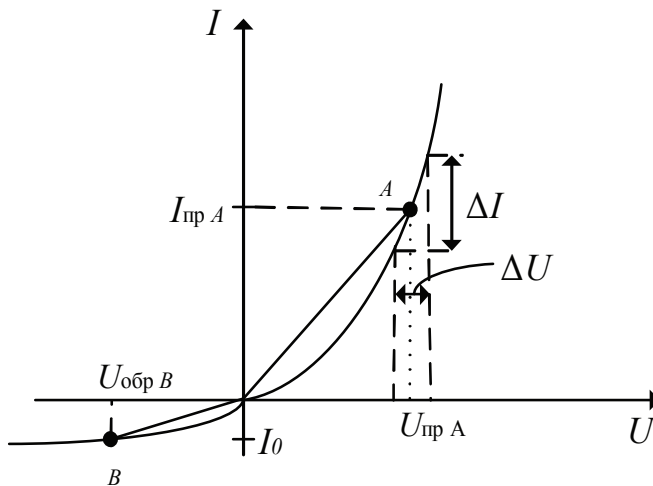


Рис. 81. Определение сопротивлений перехода

Нелинейность ВАХ определила следующие соотношения между сопротивлениями: прямая ветвь $r_{\text{пер}} \ll R_{\text{пер}}$; обратная ветвь $R_{\text{пер}} \ll r_{\text{пер}} \rightarrow \infty$.

Вольт-фарадная характеристика p - n перехода

ВАХ описывает связь между I и U на p - n переходе в статическом режиме. Эти уравнения могут быть использованы для определения I перехода и при воздействии переменного U , изменяющегося с небольшой

скоростью. Но они теряют силу при быстром изменении напряжения, соизмеримом по длительности цикла с временем накопления и высасывания неравновесного носителя заряда в базе. Такой режим называют динамическим.

В динамическом режиме *p-n* переход является инерционным элементом по отношению к быстрым изменениям тока или напряжения, поскольку новое распределение носителей устанавливается не сразу. Во-первых, внешнее напряжение меняет ширину перехода, а значит, и величину пространственных зарядов в переходе, во-вторых, при инжекции и экстракции меняются заряды в квазинейтральной области базы. Следовательно, наряду с проводимостью, которая в первом приближении характеризуется выражением $I = I_0 \left(\exp[U/\phi_T] - 1 \right)$, *p-n* пе-

реход обладает емкостью, которую можно считать подключенной параллельно переходу. Эту емкость принято разделять на две составляющие: барьерную емкость $C_{\text{бар}}$, отражающую перераспределение зарядов в переходе, и диффузионную емкость $C_{\text{диф}}$, отражающую перераспределение носителей зарядов в базе.

Изменение внешнего напряжения dU на *p-n* переходе приводит к изменению накопленного в нем заряда dQ . Поэтому *p-n* переход ведет себя подобно конденсатору, емкость которого $C = dQ/dU$.

В зависимости от физической природы изменяющегося заряда различают емкости зарядную (барьерную) и диффузионную.

Барьерная емкость определяется изменением нескомпенсированного заряда ионов при изменении ширины запирающего слоя под воздействием внешнего обратного напряжения. Поэтому идеальный электронно-дырочный переход можно рассматривать как плоский конденсатор, емкость которого определяется по соотношению

$$C_{\text{бар}} = \epsilon \Pi / \delta,$$

где Π , δ — соответственно площадь и толщина *p-n* перехода. Кроме того,

$$C_{\text{бар}} = \Pi \sqrt{\frac{\epsilon q N_a N_d}{2(U_k + |U_{\text{обр}}|)(N_a + N_d)}}.$$

В общем случае зависимость зарядной емкости от приложенного к *p-n* переходу обратного напряжения выражается по формуле

$$C_{\text{бар}} = \frac{C_0}{(1 + |U_{\text{обр}}|/U_{\kappa})^{\gamma}},$$

где C_0 — емкость p - n перехода при $U_{\text{обр}} = 0$; γ — коэффициент, зависящий от типа p - n перехода (для резких p - n переходов $\gamma = 1/2$, а для плавных — $\gamma = 1/3$).

Барьерная емкость увеличивается с ростом N_a и N_d , а также с уменьшением обратного напряжения. Характер зависимости $C = f(U_{\text{обр}})$ показан на рис. 82.

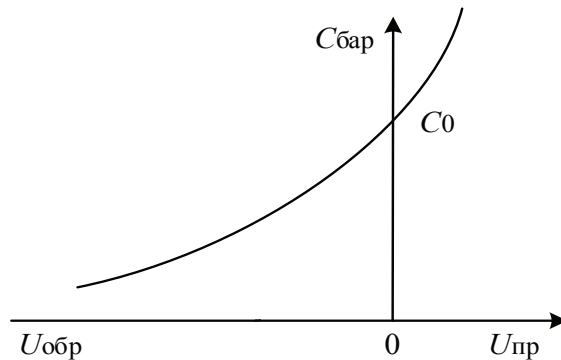


Рис. 82. Зависимость барьерной емкости от внешнего напряжения

Одна из принципиальных особенностей, отличающих емкость $C_{\text{бар}}$ от емкости плоского конденсатора, состоит в том, что в переходе направление вектора напряженности электрического поля не зависит от полярности приложенного напряжения. Оценим значение $C_{\text{бар}}$ для кремниевого p - n перехода, полагая $l_0 = 0,5$ мкм; $\phi_K = 0,75$ В; $S = 1$ мм² и $U_{\text{обр}} = 10$ В, тогда $C_{\text{бар}} \approx 10$ пФ.

Рассмотрим диффузионную емкость. При увеличении внешнего напряжения, приложенного к p - n переходу в прямом направлении, растет концентрация инжектированных носителей вблизи границ перехода, что приводит к изменению количества заряда, обусловленного неосновными носителями в p - и n -областях. Это можно рассматривать как проявление некоторой емкости. Поскольку она зависит от изменения диффузионной составляющей тока, она называется диффузионной. Диффузионная емкость представляет собой отношение приращения инжекционного заряда $dQ_{\text{инж}}$ к вызвавшему его изменению напряжения $dU_{\text{пр}}$, т. е.

$$C_{\text{диф}} = dQ_{\text{инж}}/dU_{\text{пр}}.$$

Можно определить заряд инжектированных носителей, например дырок в n -области,

$$Q_{p.\text{инж}} = q\Pi \int_{\delta_n}^{\infty} (p_n(x) - p_{n_0}) dx = qL_p \Pi p_{n_0} (\exp(U_{\text{пр}} / \varphi_T) - 1),$$

тогда диффузионная емкость, обусловленная изменением общего заряда неравновесных дырок в n -области, определится по формуле

$$C_{\text{диф},p} = \frac{dQ_p}{dU_{\text{пр}}} = \frac{qL_p \Pi p_{n_0}}{\varphi_T} \exp \frac{U_{\text{пр}}}{\varphi_T}.$$

Диффузионная емкость, обусловленная инжекцией электронов в p -область,

$$C_{\text{диф},n} = \frac{dQ_n}{dU_{\text{пр}}} = \frac{qL_n \Pi n_{p_0}}{\varphi_T} \exp \frac{U_{\text{пр}}}{\varphi_T}.$$

Общая диффузионная емкость

$$C_{\text{диф}} = C_{\text{диф},p} + C_{\text{диф},n} = (q\Pi / \varphi_T) (L_n n_{p_0} + L_p p_{n_0}) \exp(U_{\text{пр}} / \varphi_T).$$

Оценим значение $C_{\text{диф}}$ при $I_{\text{пр}} = 10$ мА и $\tau = 10^{-5}$ с, тогда $C_{\text{диф}} = 4$ мкФ.

Полная емкость p - n перехода определяется суммой барьерной и диффузионной емкостей $C_{\text{пер}} = C_{\text{бар}} + C_{\text{диф}}$. При включении p - n перехода в прямом направлении преобладает диффузионная емкость, а при включении в обратном направлении — барьерная.

Вольт-фарадная характеристика — это зависимость емкости p - n перехода от напряжения, приложенного к нему (рис. 83).

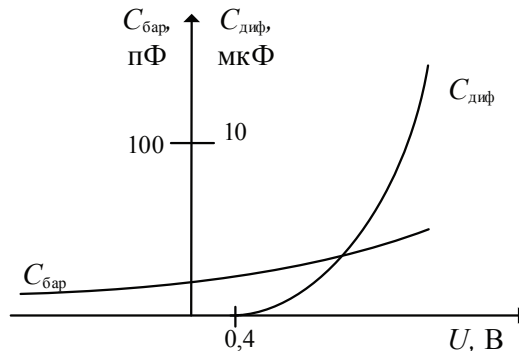


Рис. 83. Вольт-фарадная характеристика

8.9. РЕАЛЬНЫЙ p - n ПЕРЕХОД

При выводе уравнения ВАХ идеального p - n перехода учитывались лишь самые главные физические эффекты: инжекция и экстракция неосновных носителей заряда и их диффузия в нейтральных областях, прилегающих к p - n переходу. В реальных p - n переходах наблюдаются и другие физические эффекты, влияющие на ход ВАХ.

Прямая ветвь ВАХ реального p - n перехода

Под прямой ветвью ВАХ реального p - n перехода понимается зависимость прямого тока перехода от величины прямого напряжения $I_{пр} = f(U_{пр})$, которая описывается выражением

$$I_{пр} = I_0 [\exp(U_{пр}/\phi_T) - 1]$$

и должна быть экспоненциальной, как показано на рис. 84 (кривая 1; прямая ветвь ВАХ реального p - n перехода изображена кривой 2).

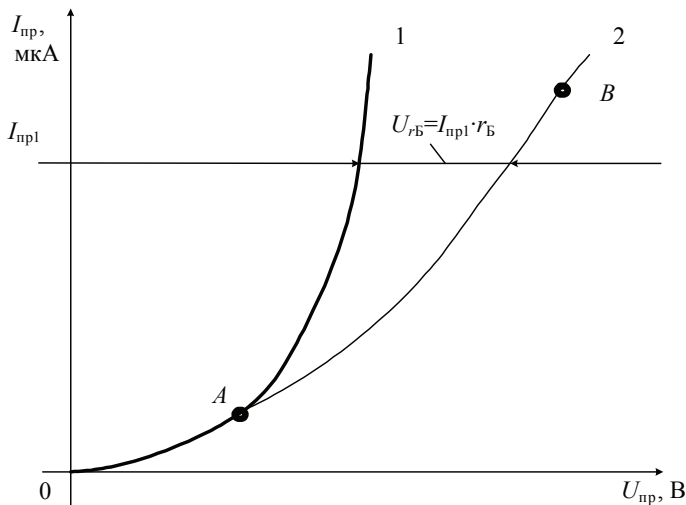


Рис. 84. Прямая ветвь ВАХ p - n перехода:

1 — идеальный p - n переход; 2 — реальный p - n переход

На прямую ветвь ВАХ реального p - n перехода оказывают влияние: материал полупроводника, используемый для изготовления p - n пере-

хода; объемное сопротивление базы *p-n* перехода; температура окружающей среды; процессы генерации и рекомбинации в самом *p-n* переходе; уровень инжекции; реальная ширина и конфигурация *p-n* перехода; характер изменения концентрации примесей; толщина прилегающих областей и т. д.

Характеристика близка к экспоненциальной только в начале зависимости — участок *OA* ВАХ, а далее рост тока при увеличении прямого напряжения замедляется и характеристика становится более пологой — участок *AB* ВАХ. Этот участок характеристики называют омическим, поскольку здесь оказывает влияние объемное сопротивление r_B слаболегированной области — базы.

Если сопротивление *p-n* перехода обозначить $r_{пер}$, то кристалл полупроводника с *p-n* переходом можно представить в виде последовательного соединения резисторов $r_{пер}$ и r_B (рис. 85).

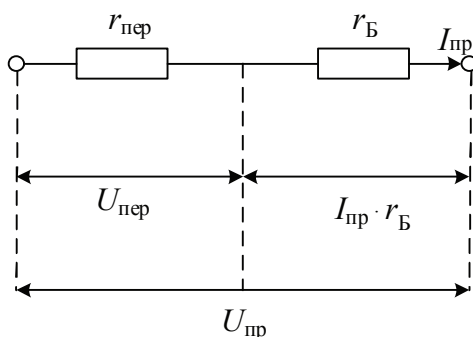


Рис. 85. Распределение внешнего напряжения

Ток, протекая через r_B , создает падение напряжения

$$U_{r_B} = I_{пр} \cdot r_B.$$

При этом внешнее напряжение не полностью падает на *p-n* переходе, а распределяется между ним и слоем базы. С учетом этого уравнение реальной ВАХ принимает вид

$$I_{пр} = I_0 \left[\exp((U_{пр} - r_B I_{пр}) / \phi_T) - 1 \right].$$

Объемное сопротивление базы находится по формуле

$$r_B = \rho_B W_B / S,$$

где ρ_B — удельное электрическое сопротивление полупроводника области базы; W_B — ширина базы; S — площадь сечения базы.

Влияние объемного сопротивления базы на прямую ветвь ВАХ реального p - n перехода проявляется в виде смещения прямой ветви в сторону больших значений прямых напряжений. Поэтому чем больше r_b , тем полнее идет прямая ветвь ВАХ реального p - n перехода, как и отмечено на рис. 85. Как правило, p - n переходы с большими значениями r_b выполняются для повышения «высоковольтности», т. е. для увеличения допустимого рабочего обратного напряжения на p - n переходе.

Даже при одинаковых условиях (одинаковая концентрация примесей, постоянная температура окружающей среды) ВАХ p - n переходов, выполненных из разных полупроводниковых материалов, различны. Главная причина этого отличия — различное значение ширины запрещенной зоны полупроводниковых материалов. Чтобы появился прямой ток, необходимо уменьшить величину потенциального барьера. Для этого на p - n переход нужно подать прямое напряжение близкое к значению контактной разности потенциалов. В p - n переходе на основе германия $\phi_K = 0,3 \dots 0,4$ В, в p - n переходе на основе кремния $\phi_K = 0,6 \dots 0,8$ В, а в p - n переходе на основе арсенида галлия $\phi_K = 1,0 \dots 1,2$ В, поэтому прямая ветвь ВАХ кремниевого p - n перехода относительно германиевого смещается вправо на $0,3$ – $0,5$ В, а в p - n переходе на основе арсенида галлия это смещение ВАХ происходит еще больше, что и отражено на рис. 86.

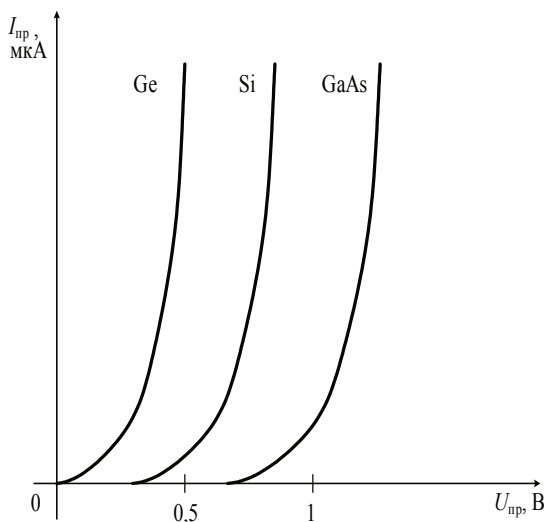


Рис. 86. Прямые ветви ВАХ p - n переходов, изготовленных из разных материалов

Сдвиг вправо прямых ветвей ВАХ кремниевых и арсенидогаллиевых *p-n* переходов объясняется из уравнения ВАХ тем, что

$$I_{0\text{Si}} < I_{0\text{Ge}}, I_{0\text{GaAs}} \ll I_{0\text{Si}},$$

следовательно, для протекания одинакового прямого тока должно выполняться соотношение

$$U_{\text{пр Si}} > U_{\text{пр Ge}} \text{ и } U_{\text{пр GaAs}} > U_{\text{пр Si}}.$$

С увеличением температуры окружающей среды растет прямой ток *p-n* перехода. Выражение для прямого тока можно записать в виде

$$I_{\text{пр}} \approx I_0 \left[\exp(U_{\text{пр}}/\phi_T) \right],$$

используя выражение для I_0 , можно записать выражение для прямого тока в виде

$$I_{\text{пр}} = \text{const} \exp \left[-\frac{(\Delta W_3 - qU_{\text{пр}})}{kT} \right],$$

откуда следует, что при увеличении температуры экспонента, прямой ток увеличивается, т. к. показатель экспоненты отрицателен $(\Delta W_3 - qU_{\text{пр}}) < 0$.

Влияние температуры на прямую ветвь ВАХ реального *p-n* перехода представлено на рис. 87.

Для оценки влияния температуры на ВАХ вводится температурный коэффициент напряжения прямой ветви, под которым понимается величина, показывающая, насколько изменится прямое напряжение для получения одной и той же величины прямого тока при изменении температуры на 1 градус.

$$\text{ТКН}_{\text{пр}} = \Delta U_{\text{пр}}/\Delta T \quad \text{при } I_{\text{пр}} = \text{const}.$$

По ВАХ, приведенным на рис. 87, температурный коэффициент напряжения прямой ветви определяется как

$$\text{ТКН}_{\text{пр}} = \Delta U_{\text{пр}}/\Delta T = (U_{\text{пр2}} - U_{\text{пр1}})/(T_2 - T_1) \text{ при } I_{\text{пр}} = I_{\text{пр1}}.$$

Обычно $\text{ТКН}_{\text{пр}} \approx -(1...3) \text{ мВ}/^\circ\text{С}$. Как видно, значение $\text{ТКН}_{\text{пр}}$ меньше нуля. Физическое объяснение этого факта сводится к следующему. При увеличении температуры уменьшается контактная разность потенциалов (а значит, и потенциальный барьер), энергия основных

носителей заряда возрастает, следовательно, возрастает количество основных носителей, энергия которых больше потенциального барьера, соответственно этому растет диффузионная составляющая тока и прямой ток увеличивается (на рис. 87, при фиксированном значении $U_{пр} = U_{пр2}$, прямой ток возрастает от значения $I_{пр2}$ до $I_{пр1}$).

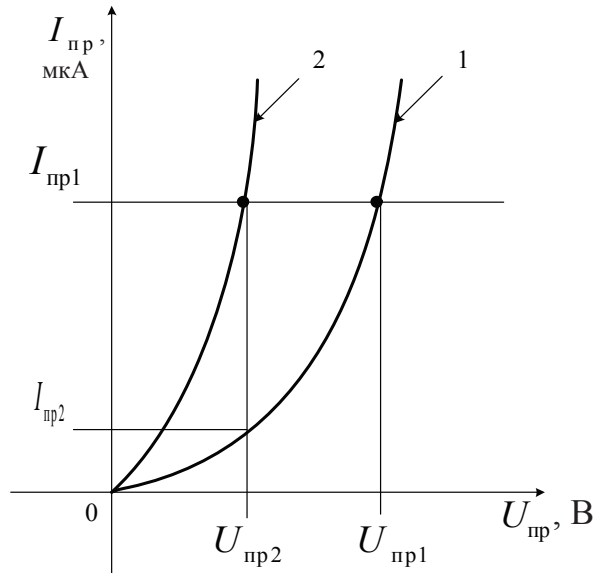


Рис. 87. Влияние температуры на прямую ветвь ВАХ p - n перехода:
1 — $T_1 = +20\text{ }^{\circ}\text{C}$; 2 — $T_2 = +50\text{ }^{\circ}\text{C}$

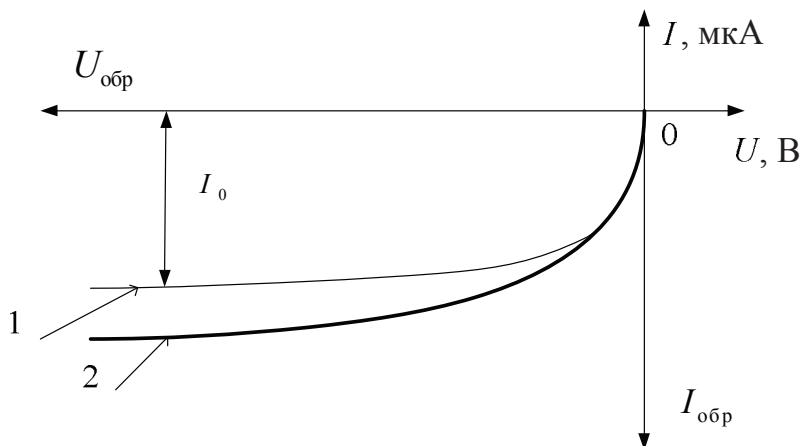
Обратная ветвь ВАХ реального p - n перехода

Под обратной ветвью вольт-амперной характеристики реального p - n перехода понимается зависимость обратного тока от значения обратного напряжения: $I_{обр} = f(U_{обр})$ (рис. 88).

Отличие реальной обратной ветви ВАХ p - n перехода от идеальной состоит в следующем: обратный ток растет при увеличении обратного напряжения p - n перехода и имеет значение большее, чем I_0 . Это объясняется тем, что в реальном p - n переходе обратный ток содержит несколько составляющих

$$I_{обр} = I_0 + I_{т/г} + I_y,$$

где I_0 — ток насыщения, или тепловой ток; $I_{т/г}$ — ток термогенерации; I_y — ток утечки.

Рис. 88. Обратная ветвь ВАХ p - n перехода:

1 — идеальный переход; 2 — реальный переход

Поясним сущность основных составляющих обратного тока реального p - n перехода.

Тепловой ток, или ток насыщения, I_0 обусловлен тепловой генерацией электронно-дырочных пар атомами собственных полупроводников в областях, примыкающих к p - n переходу на расстоянии, равном длине диффузии (рис. 89).

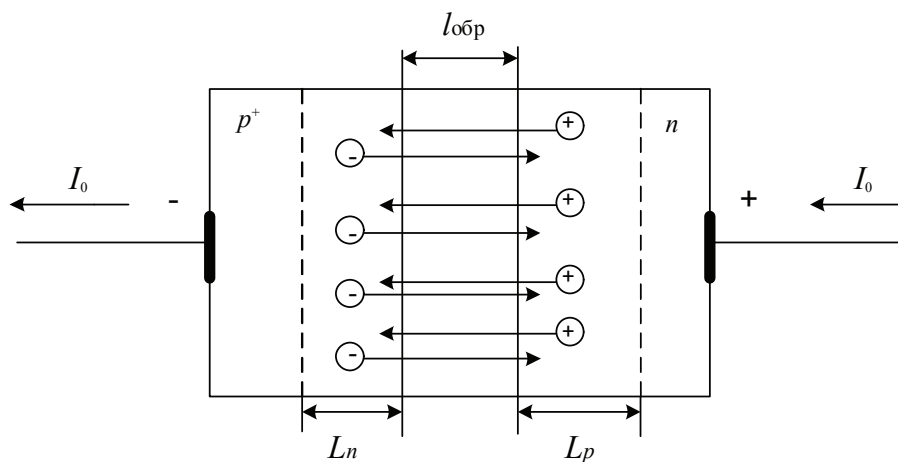


Рис. 89. Образование тока насыщения:

$l_{обр}$ — ширина обратносмещенного p - n перехода; S — площадь p - n перехода;
 L_n — объем диффузии в полупроводнике p -типа неосновных носителей заряда — электронов; L_p — объем диффузии в полупроводнике n -типа неосновных носителей заряда — дырок

Ток I_0 не зависит от величины обратного напряжения, а зависит от материала полупроводника, температуры окружающей среды, степени легирования полупроводников.

Следует отметить, что ток насыщения I_0 кремниевых p - n переходов много меньше обратного тока германиевых p - n переходов. Это связано с различием ширины запрещенной зоны: $\Delta W_{\text{Ge}} = 0,72$ эВ; $\Delta W_{\text{Si}} = 1,12$ эВ. Ток насыщения определяется неосновными носителями заряда примесного полупроводника. Так, например, в полупроводнике n -типа — это дырки p_n , которые определяются в соответствии с законом действующих масс $p_n = n_i^2/n_n \approx n_i^2/N_D$. Поскольку $n_{i\text{Ge}} \approx 10^{13} \text{ см}^{-3}$, а $n_{i\text{Si}} \approx 10^{10} \text{ см}^{-3}$, при равной концентрации примеси, концентрация неосновных носителей заряда в кремниевом полупроводнике на шесть порядков меньше, чем в германиевом примесном полупроводнике, что приводит к значительной разнице значений тока насыщения.

Влияние температуры на тепловой ток можно пояснить, используя выражение

$$I_0 = BS \exp(-\Delta W_3/kT),$$

откуда следует, что, при увеличении температуры, тепловой ток возрастает экспоненциально, т. е. увеличивается в два раза при изменении температуры на каждые десять градусов Цельсия. Например, при $T_1 = +20^\circ\text{C}$ ток $I_{01} = 10$ мкА, а при $T_2 = +50^\circ\text{C}$ он определится из соотношения

$$I_{02} = I_{01} \cdot 2^{(T_2 - T_1)/10} = 10 \cdot 2^3 = 80 \text{ мкА},$$

т. е. при изменении температуры на $\Delta T = +30^\circ\text{C}$ тепловой ток возрастает в восемь раз.

Влияние концентрации примеси в примесных полупроводниках, образующих p - n переход, прослеживается при рассмотрении закона действующих масс применительно к определению концентрации неосновных носителей заряда. С ростом концентрации примеси N_A , N_D , в p - и n -областях уменьшается концентрация неосновных носителей, что ведет к уменьшению теплового тока.

Ток термогенерации появляется из-за конечной ширины p - n перехода, не учитываемой теорией идеального p - n перехода, и обусловлен генерацией электронно-дырочных пар в объеме p - n перехода (рис. 90).

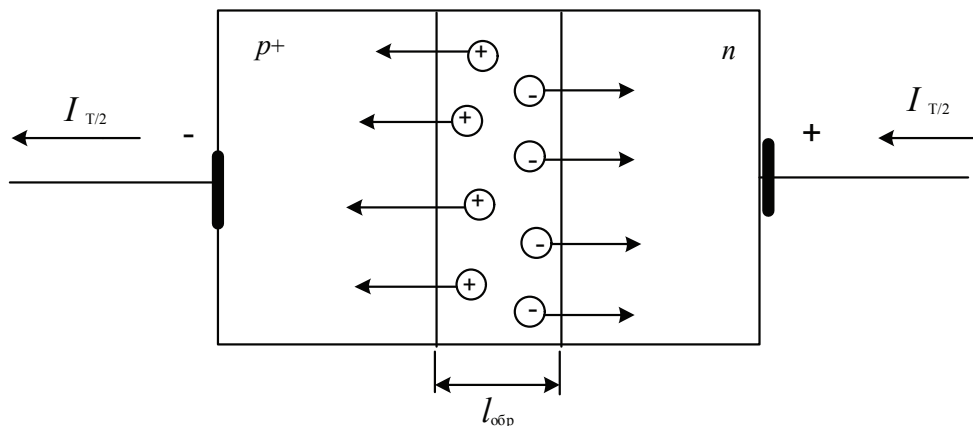


Рис. 90. Схема образования тока термогенерации

В отсутствие внешнего напряжения, между процессами генерации и рекомбинации устанавливается равновесие. При приложении к p - n переходу обратного напряжения, дырки и электроны, образующиеся в результате генерации, выводятся полем запирающего слоя. Это приводит к возникновению дополнительного тока термогенерации $I_{т/г}$, совпадающего с обратным током p - n перехода,

$$I_{т/г} = qSl_{обр} \cdot \frac{n_i}{\tau}.$$

Ток термогенерации зависит от $U_{обр}$, т. к. с увеличением обратного напряжения на p - n переходе происходит его расширение в соответствии с выражением

$$l_{обр} \approx l_0 \cdot \sqrt{\frac{|U_{обр}|}{\Phi_K}}.$$

Отсюда можно сделать вывод, что ток термогенерации пропорционален корню квадратному из значения обратного напряжения на p - n переходе. Кроме этого, $I_{т/г}$ зависит от материала полупроводника, температуры окружающей среды, концентрации примеси в p - и n -областях p - n перехода.

Как и для теплового тока, влияние материала полупроводника на ток термогенерации связано с различной шириной запрещенной зоны, а это определяет отличие в значении контактной разности потенциалов и соответственно в величине ширины p - n перехода l_0 . Поэтому имеем

для кремниевого полупроводника

$$\Delta W_3 = 1,12 \text{ эВ}, I_{\text{т/г}} \approx 10^3 I_0;$$

для германиевого полупроводника

$$\Delta W_3 = 0,72 \text{ эВ}, I_{\text{т/г}} \approx 0,1 I_0.$$

Из приведенных значений соотношения между током термогенерации и тепловым током можно отметить, что для германиевых p - n переходов ток термогенерации не учитывается, а для кремниевых p - n переходов пренебрегают током насыщения. Поэтому на ВАХ кремниевых p - n переходов нет выраженного участка насыщения.

При увеличении температуры окружающей среды, ток термогенерации возрастает экспоненциально, т. е. удваивается при изменении температуры на каждые 10–14 градусов. Механизм влияния температуры аналогичен тому, что был рассмотрен при характеристике теплового тока.

От степени легирования областей p - n перехода ток термогенерации зависит следующим образом. С ростом N_A , N_D происходит увеличение контактной разности потенциалов ϕ_K , уменьшение I_0 и $I_{\text{обр}}$, а следовательно, и уменьшение тока термогенерации.

Ток утечки I_y обусловлен поверхностными энергетическими уровнями, которые способствуют активной генерации — рекомбинации, а также проводимостью поверхностных молекулярных и ионных пленок, шунтирующих p - n переход (это могут быть молекулы окислов основного материала, молекулы газов, воды, ионы водорода и т. п.). Ток утечки слабо зависит от температуры окружающей среды и линейно возрастает при изменении обратного напряжения. Почти линейный участок зависимости $I_y(U)$ можно охарактеризовать эквивалентным сопротивлением утечки R_y . Характерной особенностью тока утечки является его временная нестабильность, которую часто называют ползучестью. Ползучесть проявляется в изменении обратного тока в течение некоторого времени после скачкообразного изменения обратного напряжения, в частности, после его включения.

Проанализировав составляющие обратного тока p - n перехода, можно принять следующее решение. Обратный ток германиевого p - n перехода включает составляющие $I_{\text{обрGe}} \approx I_0 + I_y$, а обратный ток кремниевого p - n перехода — $I_{\text{обрSi}} \approx I_{\text{т/г}} + I_y$.

На рис. 91 приведены примерные зависимости составляющих обратного тока германиевого p - n перехода и обратные ветви ВАХ для двух

температур окружающей среды. Для германиевых *p-n* переходов, обратный ток в основном определяется током насыщения и имеет величины десятки микроампер. Ток термогенерации у них мал, и им обычно пренебрегают. Незначительный наклон обратной ветви ВАХ германиевых *p-n* переходов обусловлен током утечки.

На рис. 92 приведены примерные зависимости составляющих обратного тока кремниевого *p-n* перехода и обратные ветви ВАХ для двух температур окружающей среды. Обратный ток кремниевого *p-n* перехода примерно на три — четыре порядка меньше обратного тока германиевого перехода. Объясняется это тем, что ширина запрещенной зоны у кремния больше, чем у германия, а концентрация неосновных носителей заряда оказывается на шесть порядков ниже. Поэтому ток I_0 в кремниевом *p-n* переходе пренебрежимо мал, а ток термогенерации невелик из-за малого объема *p-n* перехода, ток утечки при современной технологии изготовления *p-n* перехода имеет незначительную величину. Отсюда в целом обратный ток кремниевого *p-n* перехода имеет намного меньшее значение, чем в германиевом переходе.

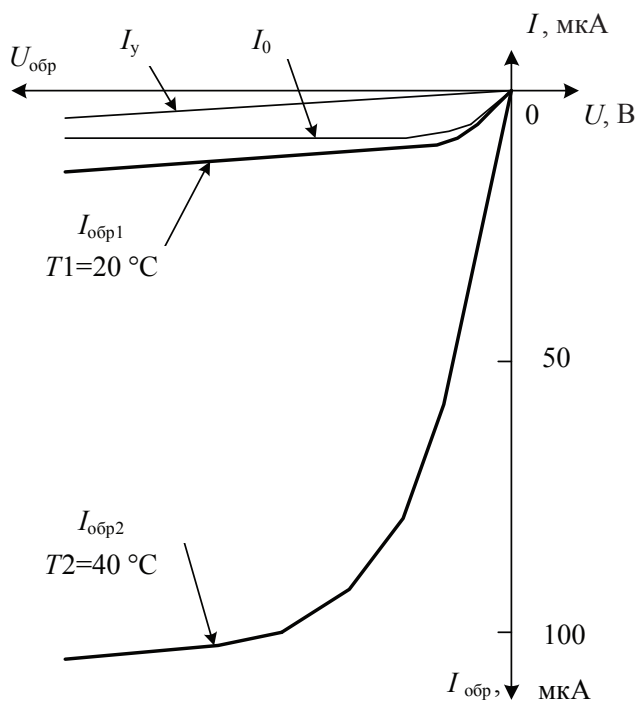
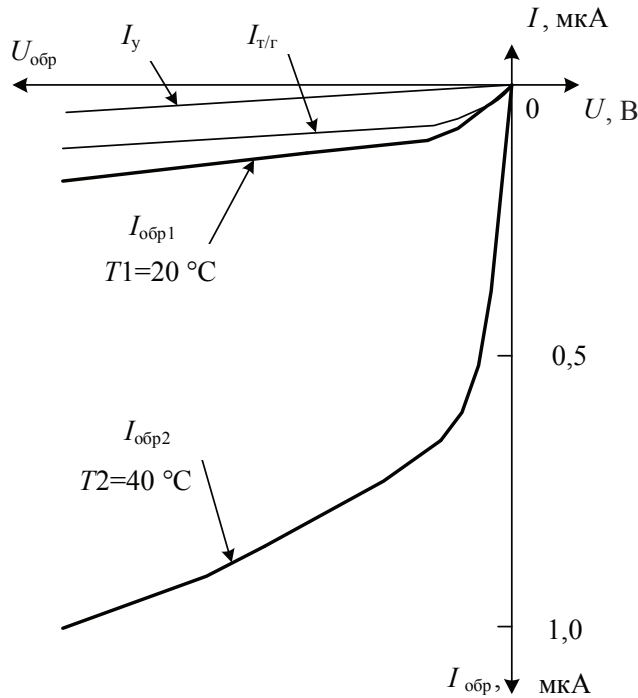


Рис. 91. Обратные ветви ВАХ германиевого *p-n* перехода

Рис. 92. Обратные ветви ВАХ кремниевого p - n перехода

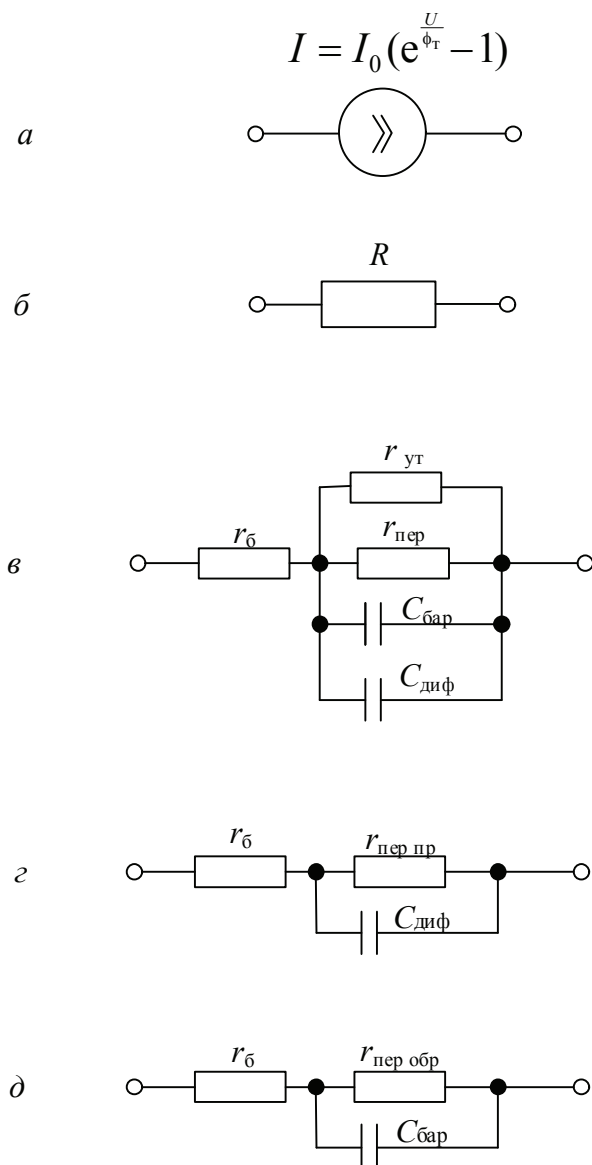
8.10. ЭКВИВАЛЕНТНЫЕ СХЕМЫ РЕАЛЬНОГО p - n ПЕРЕХОДА

Эквивалентной называется схема, составленная из идеальных элементов (индуктивность L , емкость C , сопротивления R , r , источник тока, $\text{--}(\text{--}\text{>})\text{--}$ источник напряжения $\text{--}(\text{--}\text{>})\text{--}$) и обладающая теми же свойствами, что и исходный объект в рамках решаемой задачи, но более удобная для анализа.

В зависимости от решаемой задачи, эквивалентные схемы p - n перехода могут быть различными. Некоторые из них представлены на рис. 93.

В диапазоне постоянных токов p - n переход может быть представлен источником тока с заданной ВАХ (рис. 93, а).

При фиксированном значении постоянного тока p - n переход представляет собой сопротивление постоянному току $R = U/I$, которое определяется по закону Ома (рис. 93, б).

Рис. 93. Эквивалентные схемы p - n перехода:

a — в диапазоне постоянных токов; *б* — при фиксированном значении постоянного тока; *в* — по переменному току (малосигнальная эквивалентная схема); *г* — при прямом смещении p - n перехода; *д* — при обратном смещении

Если амплитуда приложенного гармонического напряжения удовлетворяет неравенству $U_{\text{вн}} \ll \phi_T$ (режим малого сигнала), то реаль-

ный p - n переход можно описать эквивалентной схемой в виде линейного двухполюсника, это так называемая эквивалентная схема p - n перехода по переменному току, или малосигнальная эквивалентная схема (рис. 93, в). Схема содержит дифференциальное сопротивление p - n перехода $r_{\text{пер}}$, диффузионную емкость $C_{\text{диф}}$, барьерную емкость $C_{\text{бар}}$, объемное сопротивление базы $r_{\text{б}}$ (если переход является симметричным, то вместо $r_{\text{б}}$ надо учесть сопротивление объема p - и n -областей).

Характеристика инерционных свойств перехода с помощью емкостей особенно удобна в случае малых переменных сигналов, действующих на фоне больших постоянных смещений. При этом емкости оказываются почти линейными и наглядно дополняют эквивалентную схему p - n перехода.

Сопротивление утечки $r_{\text{ут}}$ учитывает реальную обратную ветвь ВАХ (ток утечки).

При прямом смещении p - n перехода следует полагать, что $C_{\text{бар}} \ll C_{\text{диф}}$, дифференциальное сопротивление $r_{\text{пер пр}}$ мало и соизмеримо с $r_{\text{б}}$, поэтому эквивалентная схема может быть упрощена (см. рис. 93, з).

При обратном смещении выполняются неравенства $C_{\text{бар}} \gg C_{\text{диф}}$, $r_{\text{пер обр}} \gg r_{\text{б}}$ (для идеальной ВАХ p - n перехода $r_{\text{пер обр}} \rightarrow \infty$) (см. рис. 93, д).

8.11. Пробой p - n перехода

Общая характеристика пробоя p - n перехода

Обратное напряжение, приложенное к диоду, падает на выпрямляющем электрическом переходе. При больших обратных напряжениях происходит пробой электрического перехода.

Пробой p - n перехода — это явление резкого уменьшения дифференциального сопротивления p - n перехода, сопровождающееся резким увеличением обратного тока, при достижении обратным напряжением критического для данного перехода значения.

Пробой приводит к выходу p - n перехода из строя лишь в том случае, когда возникает чрезмерный разогрев перехода и происходят необратимые изменения его структуры. Если же мощность, выделяющаяся в p - n переходе, не превышает максимально допустимую, переход со-

храняет работоспособность и после пробоя. Поэтому для некоторых типов диодов пробой является основным рабочим режимом.

Напряжение, при котором наступает пробой перехода, зависит от типа p - n перехода и может иметь величину от единиц до сотен вольт.

В зависимости от физических явлений, приводящих к пробую, различают тепловой, лавинный и полевой пробой. Два последних вида пробоя p - n перехода относятся к электрическому пробую. Резкий рост обратного тока p - n перехода в режиме электрического пробоя происходит за счет увеличения числа носителей заряда в переходе при достижении напряженностью поля в переходе некоторого критического значения. При тепловом пробое число носителей заряда в переходе возрастает за счет термической ионизации атомов.

Тепловой пробой p-n перехода

Тепловой пробой характерен для широких p - n переходов, у которых база слабо легирована примесями. Данный тип пробоя обусловлен разогревом p - n перехода при протекании через него обратного тока. В режиме постоянного тока мощность, выделяемая в p - n переходе, определяется по соотношению

$$P_{\text{выд}} = I_{\text{обр}} \cdot U_{\text{обр}}.$$

Отводимая от p - n перехода мощность в результате теплопроводности и дальнейшего рассеяния теплоты в окружающую среду пропорциональна перегреву p - n перехода ($T_{\text{п}} - T_{\text{окр}}$) и обратно пропорциональна тепловому сопротивлению конструкции диода R_T

$$P_{\text{отв}} = \frac{T_{\text{п}} - T_{\text{окр}}}{R_T}.$$

Тепловое сопротивление диода R_T выражается в градусах на ватт и определяет перепад температуры, необходимый для отвода 1 Вт мощности от p - n перехода в окружающую среду. Тепловое сопротивление тем меньше, чем больше теплопроводность материала K_T , площадь Π_T и чем меньше толщина слоя δ_T , проводящего тепло,

$$R_T = \frac{\delta_T}{K_T \Pi_T}.$$

Коэффициент теплопроводности K_T германия равен $0,52 \text{ Вт}/(\text{см} \cdot ^\circ\text{C})$, т.е. примерно в 7 раз меньше, чем меди, отличающейся хорошей теплопроводностью. Теплопроводность кремния выше: $K_T = 2,19 \text{ Вт}/(\text{см} \cdot ^\circ\text{C})$.

В установившемся режиме мощность, выделяющаяся на p - n переходе, и мощность, отводимая от него, должны быть равны,

$$P_{\text{выд}} = P_{\text{отв}}.$$

Если количество тепла, выделяемого на p - n переходе, превышает количество тепла, отводимого от p - n перехода, то температура перехода начинает расти. Соответственно растет обратный ток, увеличение которого определяется тепловой генерацией электронно-дырочных пар атомами полупроводников в областях, прилегающих к p - n переходу, на расстоянии длины диффузии, как указано на рис. 94. Это приводит к дальнейшему росту $P_{\text{выд}}$ и температуры перехода $T_{\text{п}}$; тепловой режим перехода теряет устойчивость, температура и ток перехода неограниченно растут, возникает тепловой пробой.

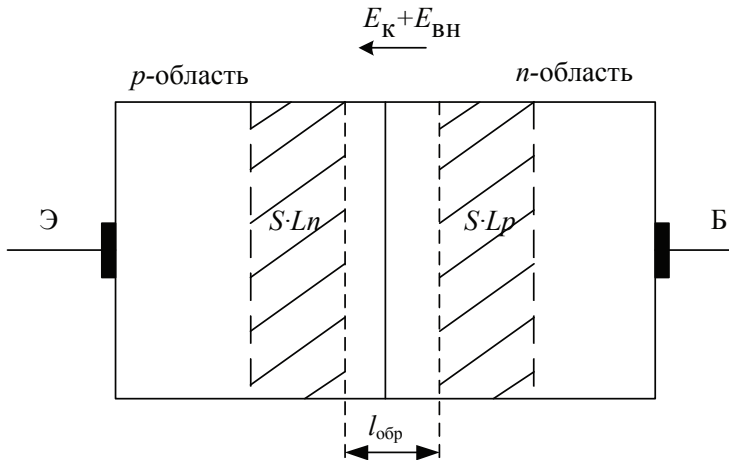
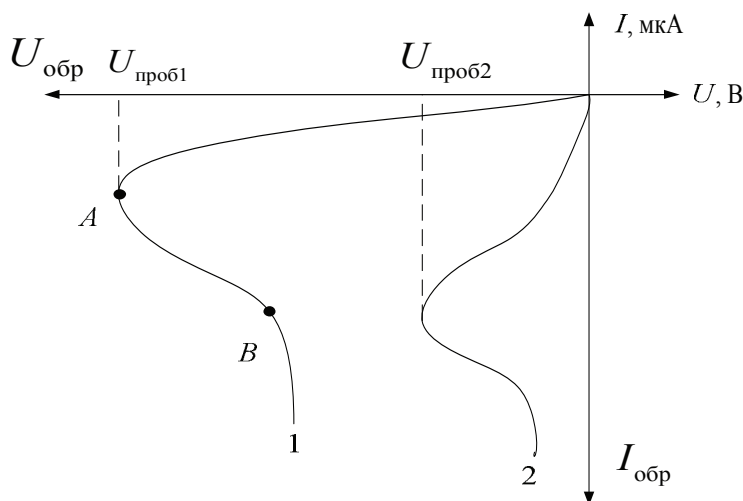


Рис. 94. Структура p - n перехода:

$S \cdot L_n$ — объем диффузии неосновных носителей — электронов в полупроводнике p -типа (S — площадь полупроводника, прилегающая к p - n переходу; L_n — длина диффузии электронов — неосновных носителей заряда полупроводника p -типа); $S \cdot L_p$ — объем диффузии неосновных носителей заряда — дырок в полупроводнике n -типа (L_p — длина диффузии дырок)

Вид обратной ветви вольт-амперной характеристики p - n перехода с тепловым пробоем представлен на рис. 95.

Рис. 95. Обратная ветвь ВАХ p - n перехода с тепловым пробоем

В точке A обратное напряжение на p - n переходе достигает значения напряжения теплового пробоя $U_{\text{проб1}}$, при котором начинается быстрый рост $I_{\text{обр}}$. Ветвь ВАХ p - n перехода с тепловым пробоем имеет участок AB , на котором дифференциальное сопротивление отрицательно

$$r_{\text{диф}} = dU_{\text{обр}}/dI_{\text{обр}} < 0,$$

т. к. концентрация носителей заряда резко увеличивается и электрическое сопротивление перехода уменьшается относительно быстрее, чем растет ток перехода.

Зависимость 1 на рис. 95 приведена для температуры окружающей среды $T_1 = +20$ °C, тепловой пробой наступает при напряжении равном $U_{\text{проб1}}$. Если температура окружающей среды возрастет до значения $T_2 = +40$ °C, то обратная ветвь ВАХ p - n перехода принимает вид зависимости 2 на рис. 95. Известно, что с увеличением температуры тепловой ток возрастает экспоненциально (т. к. резко возрастает количество неосновных носителей заряда)

$$I_0 = BS \exp [-\Delta W_3/kT],$$

где B — коэффициент, зависящий от типа полупроводника и p - n перехода; S — площадь p - n перехода; ΔW_3 — ширина запрещенной зоны полупроводника; k — постоянная Больцмана, $k = 1,38 \cdot 10^{-23}$ Дж/°C; T — температура.

По этой причине при $T = T_2$ тепловой пробой наступает раньше: при меньшем напряжении пробоя равном $U_{\text{проб}2}$. Пробивное напряжение уменьшается, во-первых, в связи с увеличением выделяющейся мощности при тех же обратных напряжениях и, во-вторых, из-за ухудшения теплоотвода от p - n перехода. Это означает, что температурный коэффициент напряжения для теплового пробоя имеет отрицательное значение

$$\text{ТКН}_{\text{ТЕПЛ}} = \Delta U_{\text{проб}} / \Delta T < 0,$$

где $\Delta U_{\text{проб}}$ — изменение напряжения пробоя при изменении температуры на величину $\Delta T = T_2 - T_1$ при фиксированном значении обратного тока, $\Delta U_{\text{проб}} = U_{\text{проб}2} - U_{\text{проб}1}$.

Тепловой пробой — необратимый пробой, поскольку может привести к плавлению полупроводникового материала. В связи с тем, что пробивное напряжение при тепловом пробое зависит от обратного тока через p - n переход, в диодах с большими обратными токами даже при комнатных температурах создаются условия для теплового пробоя и он наступает раньше, чем лавинный пробой. Это справедливо для германиевых полупроводниковых диодов и мощных транзисторов. И наоборот, в кремниевых диодах, из-за значительно меньших обратных токов, напряжение теплового пробоя получается настолько большим, что раньше наступает лавинный пробой. Однако это не означает, что в кремниевых диодах не может быть теплового пробоя. Он может происходить при высоких температурах окружающей среды. Кроме того, пробой может начаться как лавинный, а затем, по мере увеличения обратного тока, перейти в тепловой пробой.

В связи с тем, что напряжение пробоя уменьшается с увеличением теплового сопротивления, следует обратить особое внимание на совершенство конструкции диода. Тепловой пробой наблюдается и тогда, когда имеет место плохой отвод тепла от корпуса полупроводникового прибора.

Полевой пробой

Полевой, или туннельный, пробой относится к электрическому виду пробоя и характерен для сравнительно узких p - n переходов (ширина p - n перехода в равновесном состоянии составляет сотые доли микрометра). Он возникает в том случае, когда обе области p - n перехо-

да имеют высокую степень легирования примесями. При этом длина свободного пробега λ электронов меньше ширины обратносмещенного p - n перехода

$$\lambda < l_{\text{обр.}}$$

При напряженности электрического поля $E = U_{\text{обр.}}/l_{\text{обр.}}$ в p - n переходе, равной критическому значению $E_{\text{кр}} = (2...4) \cdot 10^5$ В/см, происходит полевой, или туннельный, пробой. При такой большой напряженности электрического поля, у атома полупроводника происходит отрыв валентных электронов и число носителей заряда растет. С энергетической (зонной) точки зрения основу полевого пробоя составляет туннельный эффект — явление «просачивания» электронов сквозь узкий энергетический барьер p - n перехода, т. е. переход электронов с занятых энергетических уровней валентной зоны полупроводника p -типа на свободные энергетические уровни зоны проводимости полупроводника n -типа. Эти переходы происходят без изменения энергии электрона (рис. 96). На энергетической диаграмме переходы происходят на одном энергетическом уровне, т. е. горизонтально.

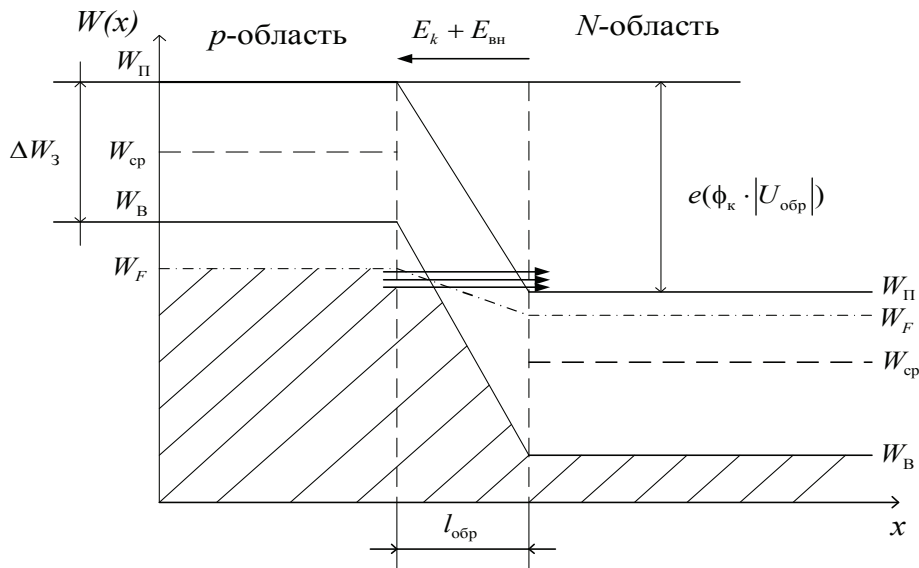


Рис. 96. Энергетическая диаграмма p - n перехода при полевом пробое

Вероятность туннельных переходов при напряженности электрического поля $E = 10^5$ В/см составляет один электрон в секунду, а при

напряженности электрического поля $E = 10^6$ В/см — 10^{12} электронов в секунду. Поэтому, при критическом значении напряженности электрического поля обратносмещенного p - n перехода, количество туннельных переходов будет значительным, что приводит к резкому увеличению обратного тока.

При дальнейшем увеличении обратного напряжения на p - n переходе $U_{\text{обр}} > U_{\text{проб}}$, рост обратного тока происходит по экспоненциальному закону. Это объясняется увеличением напряженности электрического поля и степени перекрытия валентной зоны полупроводника p -типа и зоны проводимости полупроводника n -типа.

Обратная ветвь ВАХ p - n перехода для случая полевого пробоя представлена на рис. 97. Полевой пробой имеет место в p - n переходах с напряжением пробоя до 5 В.

Зависимость 1 на рис. 98 соответствует значению температуры окружающей среды $T_1 = +20$ °С. При увеличении температуры окружающей среды до значения $T_2 = +50$ °С, ВАХ p - n перехода видоизменяется (зависимость 2). При увеличении температуры, обратный ток p - n перехода возрастает в связи с ростом концентрации неосновных носителей заряда по экспоненциальному закону. Такое изменение обратного тока наблюдается при регулировании обратного напряжения в диапазоне от нуля до напряжения пробоя.

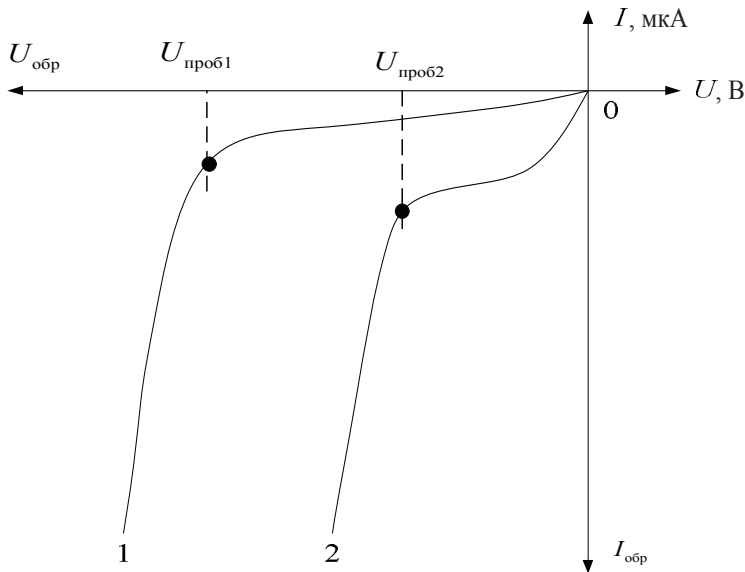


Рис. 97. Обратная ветвь ВАХ p - n перехода при полеводном пробое

С увеличением температуры напряжение пробоя уменьшается и становится равным $U_{\text{проб}2}$ (зависимость 2 на рис. 97). Это обусловлено тем, что при увеличении температуры, возрастает амплитуда тепловых колебаний атомов полупроводника в узлах кристаллической решетки, энергия электронов также растет, величина контактной разности потенциалов *p-n* перехода ϕ_K снижается, ширина *p-n* перехода $l_{\text{обр}}$ уменьшается, а напряженность электрического поля в *p-n* переходе E увеличивается, критическое значение напряженности поля $E_{\text{кр}}$ достигается при меньшем значении $U_{\text{обр}}$, растет количество туннельных переходов и, следовательно, резко возрастает обратный ток. Таким образом, температурный коэффициент напряжения при полевом пробое имеет отрицательное значение

$$\text{ТКН}_{\text{пол}} = \Delta U_{\text{проб}} / \Delta T < 0,$$

где $\Delta U_{\text{проб}}$ — изменение напряжения пробоя при изменении температуры на величину $\Delta T = T_2 - T_1$ и фиксированном значении обратного тока, $\Delta U_{\text{проб}} = U_{\text{проб}2} - U_{\text{проб}1}$.

При полевом пробое, пробивное напряжение оказывается обратно пропорциональным концентрации примесей в областях, прилегающих к *p-n* переходу, или прямо пропорционально удельному сопротивлению этих областей.

Для кремниевых *p-n* переходов пробивное напряжение при полевом пробое определяется по эмпирическому соотношению

$$U_{\text{проб}} = 200\rho_n + 73\rho_p;$$

а для германиевых переходов

$$U_{\text{проб}} = 190\rho_n + 94\rho_p,$$

где ρ_n и ρ_p — удельные сопротивления *n*- и *p*-областей, прилегающих к переходу, Ом·см.

Итак, чем сильнее легированы области *p-n* перехода, тем меньше удельное сопротивление этих областей и тем меньше пробивное напряжение при полевом пробое.

Лавинный пробой

Лавинный пробой относится к электрическому виду пробоя и проявляется в *p-n* переходах средней величины, т. о., ширина *p-n* перехода достаточно большая. При увеличении значения обратного напря-

жения на p - n переходе, напряженность электрического поля E , В/см, растет $E = U_{\text{обр}}/l_{\text{обр}}$. Когда напряженность электрического поля достигает критического значения $E_{\text{кр}} = 80 \dots 120$ кВ/см, тогда создаются условия для ударной ионизации нейтральных атомов полупроводника непосредственно в p - n переходе быстрыми электронами или дырками, которые получили достаточное ускорение за счет действия напряженности электрического поля p - n перехода (рис. 98). Такие носители называются горячими, т. к. их средняя энергия существенно превышает энергию теплового движения kT_0 .

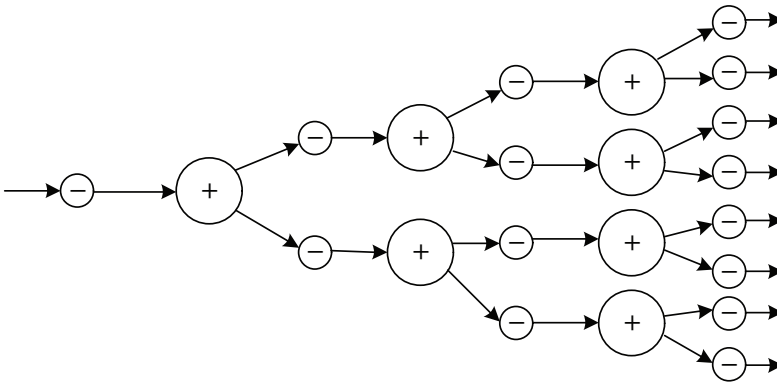


Рис. 98. Механизм ударной ионизации

В результате ударной ионизации генерируются новые пары носителей заряда, которые, в свою очередь, ускоряясь под действием напряженности электрического поля, вновь при столкновении с нейтральными атомами полупроводника образуют новые электронно-дырочные пары. Ионизацию нейтральных атомов совершают только те электроны и дырки, которые на длине свободного пробега электрона набирают за счет напряженности электрического поля энергию, достаточную для ионизации. Поэтому ширина p - n перехода должна быть достаточно большая — больше длины свободного пробега электрона $l_{\text{обр}} \gg \lambda$.

С ростом $U_{\text{обр}}$ увеличивается ширина p - n перехода и напряженность электрического поля в нем, электроны разгоняются сильнее, резко возрастает число ионизаций, совершаемых каждым электроном, и ток p - n перехода лавинообразно растет.

Напряжение лавинного пробоя определяется из соотношения

$$U_{\text{проб}} = A \rho_6^B,$$

где A, B — коэффициенты, зависящие от материала и типа электропроводности полупроводника (табл. 4); ρ_6 — удельное электрическое сопротивление базы диода.

Таблица 4

Зависимость коэффициентов A и B от материала и типа проводника

Материал и тип перехода	A	B
Германиевый переход, база p -типа	52	0,6
Германиевый переход, база n -типа	83	0,6
Кремниевый переход, база p -типа	23	0,75
Кремниевый переход, база n -типа	86	0,65

Так, например, для p - n перехода с базой n -типа

$$\rho_6 = \frac{1}{\sigma_6} \approx \frac{1}{q\mu_n n_n},$$

где q — заряд электрона; μ_n — подвижность электронов; n_n — концентрация электронов — основных носителей заряда полупроводника n -типа.

Чем меньше концентрация примесей в базе p - n перехода, тем выше ее удельное электрическое сопротивление, шире p - n переход, меньше в нем напряженность электрического поля и соответственно более высокое значение напряжения лавинного пробоя. Эмпирические коэффициенты A и B отличаются не только для p - n переходов из разных материалов, но и для переходов из одного и того же материала с разными типами электропроводности базы (p^+ - n и n^+ - p). Связано это различие коэффициентов с тем, что подвижность электронов отличается от подвижности дырок в одном и том же материале.

Лавинный пробой характерен для p - n переходов с напряжением пробоя более 7 В. Зависимость 1 на рис. 99 соответствует температуре окружающей среды $T_1 = +20^\circ\text{C}$. С увеличением температуры окружающей среды, лавинный пробой наступает при большем напряжении ($|U_{\text{проб}2}| > |U_{\text{проб}1}|$). Это объясняется тем, что с ростом температуры увеличивается амплитуда колебаний атомов кристаллической решетки полупроводника и уменьшается длина свободного пробега носителей заряда λ , а значит, и энергия, которую носитель заряда может приобрести на длине свободного пробега в электрическом поле. Поэтому для по-

лучения энергии, необходимой для ударной ионизации нейтральных атомов, требуется бóльшая напряженность электрического поля в p - n переходе, и следовательно, напряжение лавинного пробоя возрастает.

Однако при увеличении температуры, уменьшается подвижность носителей заряда полупроводника, растет удельное электрическое сопротивление базы p - n перехода и в соответствии с соотношением

$$U_{\text{проб}} \sim \rho_b$$

напряжение лавинного пробоя также возрастает.

На рис. 99 зависимость 2 изображена для температуры окружающей среды $T_2 = +50^\circ\text{C}$. В таком случае температурный коэффициент напряжения при лавинном пробое имеет положительное значение

$$\text{ТКН}_{\text{ЛАВ}} = \Delta U_{\text{проб}} / \Delta T > 0,$$

где $\Delta U_{\text{проб}}$ — изменение напряжения пробоя при изменении температуры на величину ΔT при фиксированном значении обратного тока, $\Delta U_{\text{проб}} = U_{\text{проб}2} - U_{\text{проб}1}$.

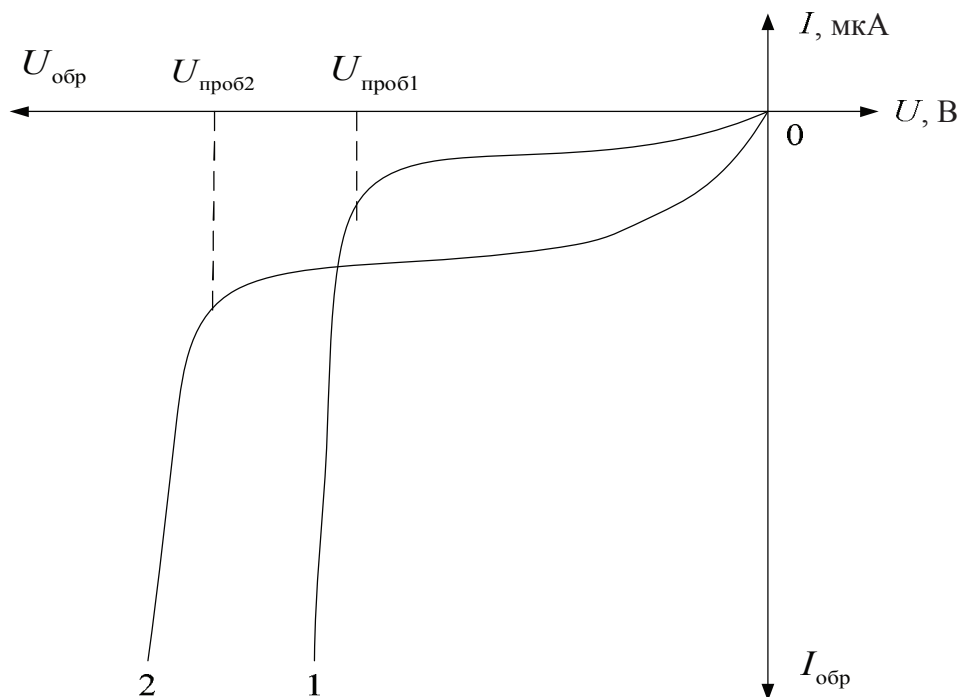


Рис. 99. Обратная ветвь ВАХ p - n перехода при лавинном пробое

8.12. ТУННЕЛИРОВАНИЕ В СИЛЬНОЛЕГИРОВАННЫХ *p-n* ПЕРЕХОДАХ

Понятие и особенности электронно-дырочных переходов с туннельным эффектом

Туннельный эффект, или туннелирование, — это преодоление микрочастицей потенциального барьера в случае, когда ее полная энергия меньше высоты барьера.

Туннельный эффект — явление исключительно квантовой природы, невозможное в классической механике. При характеристике туннельного эффекта вводится понятие вероятности прохождения частицей сквозь потенциальный барьер. Эта вероятность тем больше, чем меньше масса частицы, чем уже потенциальный барьер и чем ближе энергия частиц к значению энергии потенциального барьера, причем туннельные переходы происходят без изменения энергии частиц.

Что касается электронно-дырочных переходов, туннельный эффект наблюдается только при очень малой ширине *p-n* перехода $l_0 \leq 10^{-6}$ см, т. е. в переходах между высоколегированными p^+ - и n^+ -областями ($N_d, N_A > 10^{18}$ см $^{-3}$).

Туннельный эффект проявляется в просачивании электронов сквозь узкий энергетический (потенциальный) барьер *p-n* перехода без изменения энергии. Причем ширина энергетического барьера меньше толщины обедненного слоя полупроводника l_0 .

Вероятность туннельных переходов в электронно-дырочном переходе зависит от напряженности электрического поля E и выражается количеством переходов электронов в единицу времени, электронов/с,

$$p = \frac{N_{\text{эл}}}{t},$$

где $N_{\text{эл}}$ — количество электронов; t — время.

При напряженности электрического поля $E = 10^5$ В/см, значение $p = 1$ эл/с (один электрон в секунду), а при $E = 10^6$ В/см — $p = 10^{12}$ эл/с.

При внесении большой концентрации примесей $N_A, N_d = (10^{18} \dots 10^{20})$ см $^{-3}$, в области полупроводников *p*- и *n*-типов происходит расщепление примесных энергетических уровней с образованием примесных энергетических зон, которые проникают в разрешенные зоны: зону проводимости полупроводника *n*-типа; валентную зону по-

лупроводника p -типа. Энергетический уровень Ферми в этом случае располагается в разрешенных зонах «вырожденных» полупроводников.

Для энергетической диаграммы электронно-дырочного перехода (рис. 100) имеем

$$l_0 = \left[\frac{2\epsilon\epsilon_0\phi_k}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) \right]^{\frac{1}{2}} \cong 10^{-6} \text{ см} = 10 \text{ нм},$$

$$\Delta W_3 = q \cdot \Delta\phi_3,$$

$$\phi_k \geq \Delta\phi_3.$$

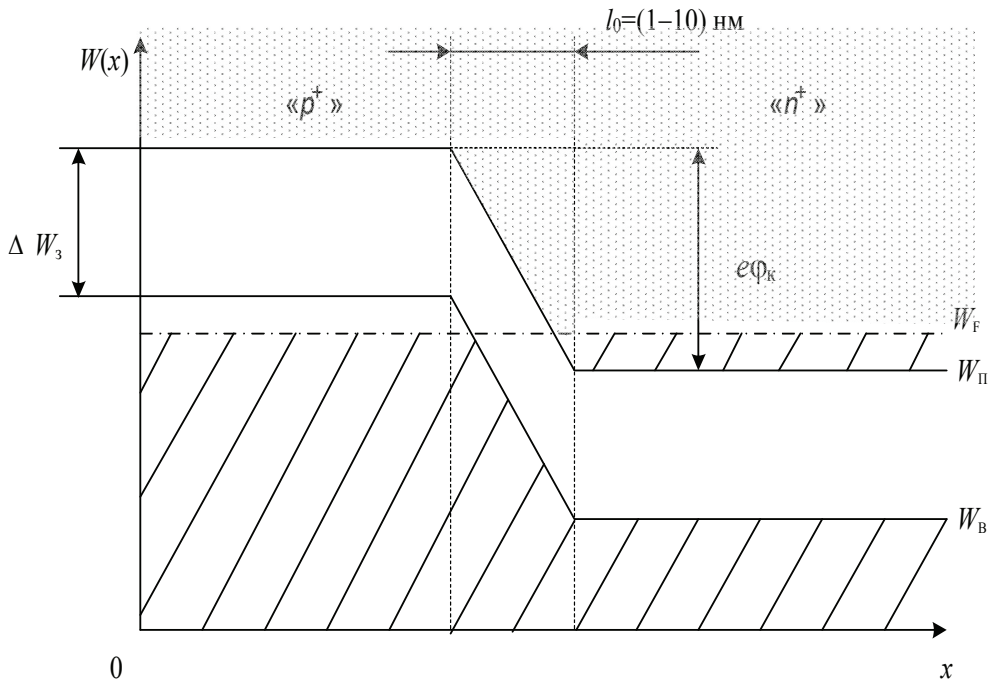



Рис. 100. Энергетическая диаграмма электронно-дырочного перехода с туннельным эффектом в равновесном состоянии:

W_B — энергетический уровень потолка валентной зоны; $W_П$ — энергетический уровень дна зоны проводимости; W_F — энергетический уровень Ферми; заштрихованы все энергетические уровни, занятые электронами;  — обозначены свободные энергетические уровни зоны проводимости

Например, для *p-n* перехода на основе германия: $\varphi_k = 0,65$ В; $l_0 = 10^{-6}$ см; $E_0 = \frac{\varphi_k}{l_0} = \frac{0,65}{10^{-6}} = 0,65 \cdot 10^6$ В/см. В таком случае в электронно-

дырочном переходе вероятности туннельных переходов электронов из *p*-области в *n*-область и обратно одинаковы. Поэтому встречные потоки электронов равны, а суммарный туннельный ток через *p-n* переход равен нулю

$$I_T = I_{Tp} + I_{Tn} = 0,$$

где I_{Tp} — туннельный ток, образованный туннельными переходами электронов из полупроводника *p*-типа в полупроводник *n*-типа; I_{Tn} — туннельный ток, создаваемый туннельными переходами электронов из полупроводника *n*-типа в полупроводник *p*-типа.

В условиях равновесного состояния электронно-дырочного перехода, изоэнергетические уровни (уровни с одинаковыми значениями энергии) по обе стороны перехода либо заняты, либо свободны с одинаковой вероятностью и туннельный ток в *p-n* переходе равен нулю ($I_T = 0$).

В электронно-дырочном переходе, образованном сильнолегированными («вырожденными») примесными полупроводниками, имеют место диффузионные и дрейфовые составляющие токов, как и в обычном *p-n* переходе,

$$I_D = I_{Dp} + I_{Dn}, \quad I_E = I_{Ep} + I_{En},$$

где I_{Dp} — дырочная составляющая диффузионного тока; I_{Dn} — электронная составляющая диффузионного тока; I_{Ep} — дырочная составляющая дрейфового тока; I_{En} — электронная составляющая дрейфового тока.

При этом концентрация основных носителей заряда в области *p-n* перехода очень высока

$$p_p \cong N_A \gg n_p, \quad n_n \cong N_D \gg p_n.$$

Диффузионная составляющая тока образуется движением основных носителей заряда в *p-n* переходе, и она будет небольшой из-за большого значения величины потенциального барьера, определяемой по контактной разности потенциалов

$$\varphi_k = \varphi_T \ln \frac{N_A N_D}{n_i^2},$$

где ϕ_k — контактная разность потенциалов; ϕ_T — температурный потенциал, $\phi_T = \frac{kT}{q}$; N_A — концентрация акцепторов в полупроводнике

p -типа; N_D — концентрация доноров в полупроводнике n -типа; n_i — концентрация носителей заряда в собственном полупроводнике. Для сильнолегированных примесных полупроводников имеет место большое значение N_A , N_D , поэтому в p - n переходе на основе данных примесных полупроводников получают высокое значение контактной разности потенциалов ϕ_k . Значение дрейфовой составляющей тока p - n перехода в данном случае невелико, т. к. мала концентрация неосновных носителей заряда в обеих областях p - n перехода. Условие равновесия электронно-дырочного перехода в общем виде запишется

$$I_D + I_E + I_T = 0,$$

или

$$I_{Dn} + I_{Dp} + I_{En} + I_{Ep} + I_{Tn} + I_{Tp} = 0.$$

Вольт-амперная характеристика идеального электронно-дырочного перехода с туннельным эффектом

При подаче внешнего напряжения на электронно-дырочный переход, принцип работы которого основан на явлении туннельного эффекта, получают вольт-амперную характеристику.

Рассмотрим вольт-амперную характеристику идеального p - n перехода при наличии туннельного эффекта (рис. 101) с помощью энергетических (зонных) диаграмм. Для точки 0 ВАХ, при условии равновесия электронно-дырочного перехода, энергетическая диаграмма изображена на рис. 100. Ее анализ показал, что полный ток через электронно-дырочный переход в этом случае равен нулю.

Для точки А (рис. 101) при небольшом значении прямого напряжения $U_{\Pi} > U_{\Pi 1} > 0$ в p - n переходе происходит уменьшение высоты потенциального барьера и смещение энергетических уровней n -области относительно энергетических уровней p -области (рис. 102).

Как показано на рис. 102, при $U_{\Pi 1} > 0$ свободные энергетические уровни валентной зоны p -области, расположенные непосредственно над энергетическим уровнем Ферми, оказываются на одной высоте с энергетическими уровнями зоны проводимости n -области, которые заполнены электронами. Поэтому будет происходить преимущественное туннели-

рование электронов из полупроводника n -типа в полупроводник p -типа на свободные энергетические уровни валентной зоны: $I_{Tn} \gg I_{Tp}$, $I \cong I_{Tn}$.

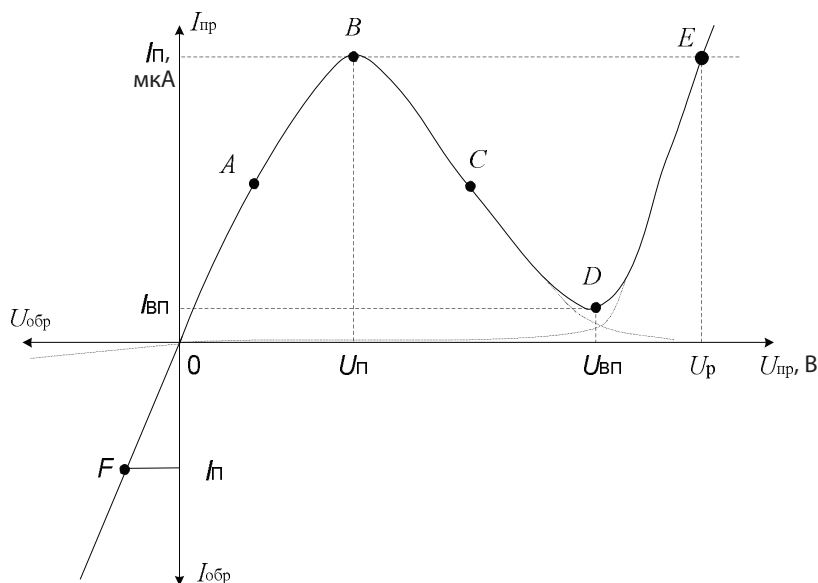


Рис. 101. Вольт-амперная характеристика идеального электронно-дырочного перехода с туннельным эффектом (----- — ВАХ электронно-дырочного перехода без учета туннельной составляющей)

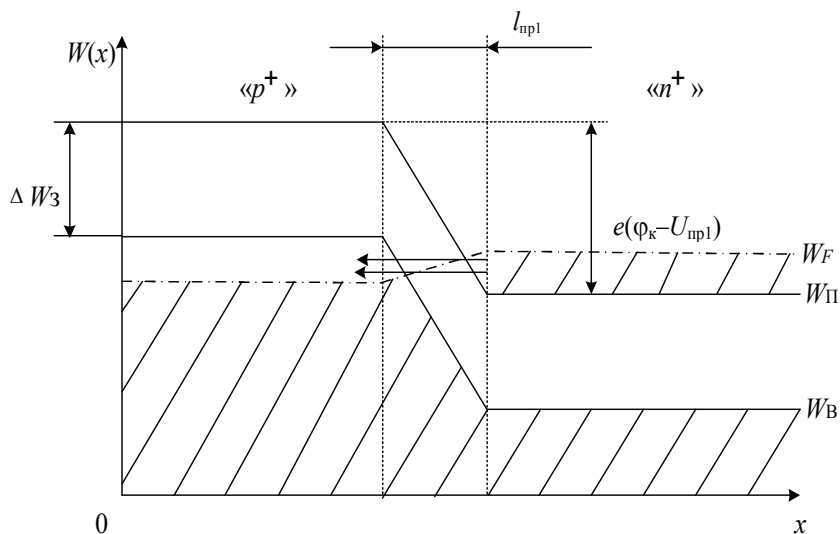


Рис. 102. Энергетическая диаграмма электронно-дырочного перехода с туннельным эффектом при $U_{\Pi} > U_{\text{пр1}} > 0$

В точке A ВАХ (см. рис. 102) имеет место и диффузионная составляющая тока p - n перехода $I_D = I_{Dn} + I_{Dp}$. Поскольку степень легирования области полупроводника n - и p -типов одинакова, токи I_{Dn} , I_{Dp} будут иметь равные значения, но их значение при данном прямом напряжении $U_{пр1}$ невелико, потому что существует еще большее значение высоты потенциального барьера ($\phi_k - U_{пр1}$) и малое количество основных носителей заряда p - и n -областей полупроводника, способных преодолеть этот потенциальный (энергетический) барьер.

Энергетическая (зонная) диаграмма для точки B ВАХ приведена на рис. 103. В этом случае прямое напряжение на p - n переходе имеет такое значение тогда, когда свободные энергетические уровни валентной и примесной зон полупроводника p -типа окажутся на одной высоте с занятыми электронами энергетическими уровнями зоны проводимости и примесной зоны полупроводника n -типа.

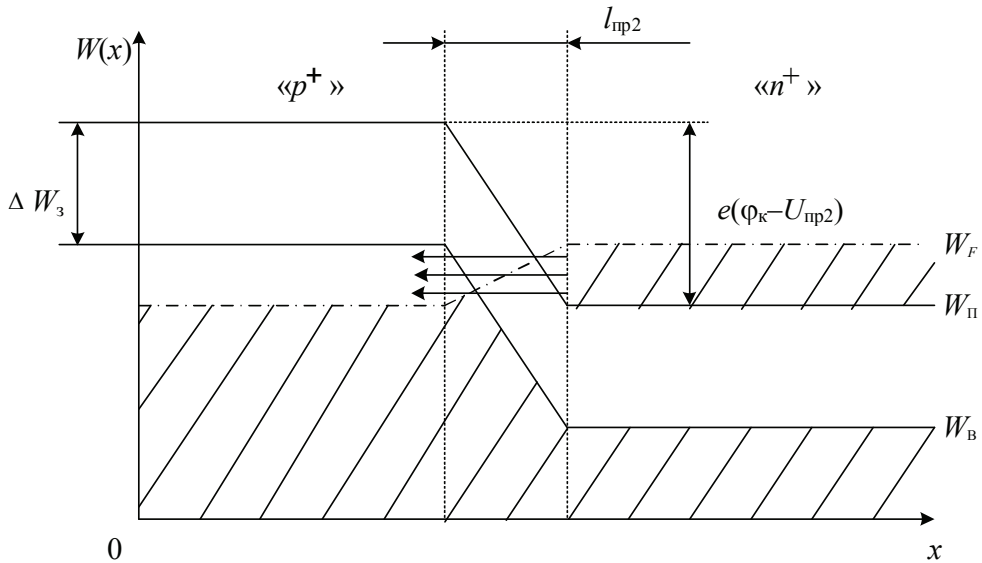


Рис. 103. Энергетическая диаграмма электронно-дырочного перехода с туннельным эффектом при $U_{пр2} = U_{П}$

В данном случае энергетический уровень Ферми полупроводника n -типа находится на одном уровне с энергией потолка валентной зоны полупроводника p -типа ($W_{Fn} = W_{Bp}$), а энергетический уровень Ферми полупроводника p -типа совпадает с энергией дна зоны проводимости

полупроводника n -типа ($W_{Fp} = W_{Fn}$). При этом перекрытие зоны проводимости полупроводника n -типа с валентной зоной полупроводника p -типа будет максимальным и количество туннельных переходов из зоны проводимости полупроводника n -типа в валентную зону полупроводника p -типа будет максимальным, значение туннельного тока p - n перехода будет также наибольшим. Отсюда при прямом напряжении $U_{np2} = U_{\Pi}$, называемым напряжением пика, ток p - n перехода $I_{\Pi} \cong I_{Tn} = \max$. Он называется током пика (см. рис. 101). В точке B ВАХ, диффузионная составляющая тока p - n перехода $I_D = I_{Dn} + I_{Dp}$ увеличивается, но по сравнению с $I_{Tn} = I_{\Pi}$, еще имеет очень малое значение и полный ток p - n перехода в основном определяется туннельной составляющей тока $I_{Tn} \gg I_D \gg I_E$, $I_{np2} < I_{np1}$.

Энергетическая (зонная) диаграмма для точки C ВАХ представлена на рис. 104.

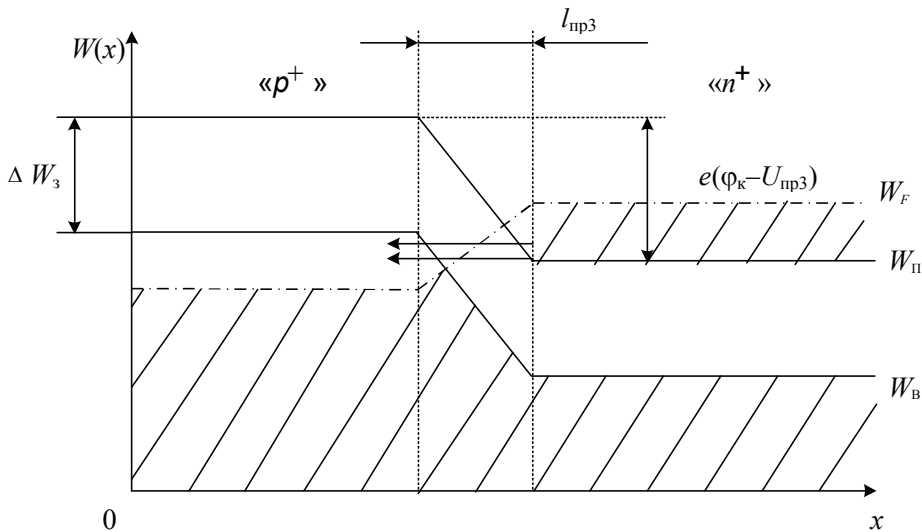


Рис. 104. Энергетическая диаграмма электронно-дырочного перехода с туннельным эффектом при $U_{np3} > U_{\Pi}$

При прямом напряжении $U_{np3} > U_{\Pi}$ происходит дальнейшее смещение энергетических уровней области полупроводника n -типа по отношению к энергетическим уровням полупроводника p -типа, перекрытие зоны проводимости полупроводника n -типа и валентной зоны полупроводника p -типа будет уменьшаться. Таким образом, количество свободных энергетических уровней валентной зоны полупроводни-

ка p -типа и количество занятых электронами энергетических уровней зоны проводимости полупроводника n -типа, находящихся напротив друг друга, будет уменьшаться, а это приводит к снижению количества туннельных переходов из зоны проводимости полупроводника n -типа в валентную зону полупроводника p -типа и, следовательно, туннельного тока p - n перехода.

Диффузионная составляющая тока в точке C ВАХ будет больше, чем в точке B (см. рис. 101), но ее удельный вес в полном прямом токе p - n перехода все еще маленький, поэтому в точке C прямой ток будет $I_{пр} \cong I_{Tn(C)} \gg I_{D(C)} \gg I_E$.

Энергетическая (зонная) диаграмма для точки D вольт-амперной характеристики изображена на рис. 105.

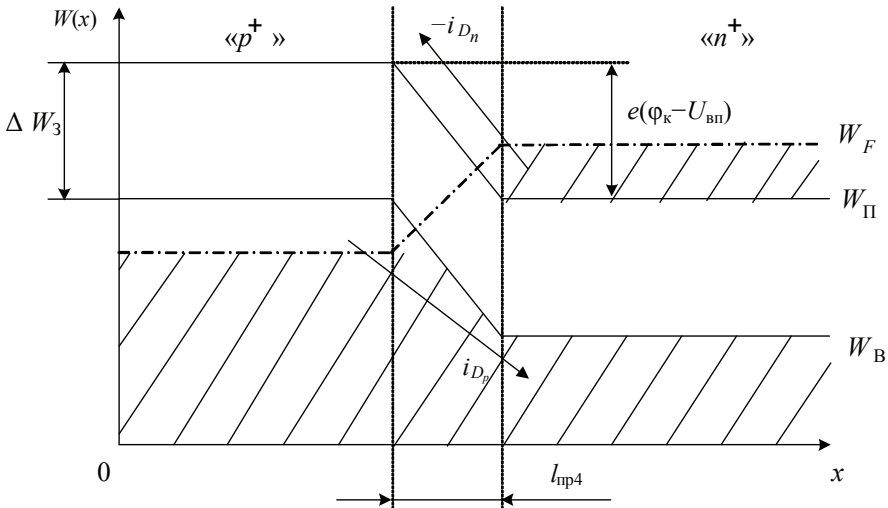


Рис. 105. Энергетическая диаграмма электронно-дырочного перехода с туннельным эффектом при $U_{пр4} = U_{ВП}$

При дальнейшем увеличении прямого напряжения $U_{пр4} > U_{пр3}$ продолжается смещение энергетических уровней полупроводника n -типа относительно энергетических уровней полупроводника p -типа. При этом свободные энергетические уровни валентной зоны полупроводника p -типа находятся напротив запрещенной зоны полупроводника n -типа, а занятые энергетические уровни зоны проводимости полупроводника n -типа находятся напротив запрещенной зоны полупроводника p -типа и условия для туннельных переходов заметно ухудшаются.

В идеальном случае туннельный эффект прекращается и $I_T = 0$. Напряжение $U_{\text{пр}4} = U_{\text{ВП}}$ получило название напряжения впадины, а прямой ток электронно-дырочного перехода, соответствующий напряжению впадины, называется током впадины $I_{\text{ВП}}$. Ток $I_{\text{ВП}}$ в точке D ВАХ (см. рис. 101) определяется по сумме избыточного туннельного тока и нарастающего тока диффузии при прямом смещении p - n перехода.

В случае возрастания прямого напряжения $U_{\text{пр}} > U_{\text{пр}4}$, туннельная составляющая тока устремляется к нулю, а диффузионная составляющая прямого тока начинает экспоненциально возрастать за счет снижения энергетического барьера p - n перехода. В точке E ВАХ, когда $I_{\text{пр}} = I_{\text{П}}$, напряжение на p - n переходе является напряжением раствора ($U_{\text{пр}} = U_{\text{р}}$) прямой ветви ВАХ электронно-дырочного перехода с туннельным эффектом.

Обратная ветвь ВАХ p - n перехода с туннельным эффектом (участок OF на рис. 101) определяется двумя составляющими токов электронно-дырочного перехода: I_E — дрейфовым током p - n перехода, который образуется за счет направленного движения неосновных носителей заряда в ускоряющем электрическом поле электронно-дырочного перехода; I_T — туннельной составляющей тока, образованной туннельными переходами электронов с заполненных энергетических уровней валентной зоны полупроводника p -типа на свободные энергетические зоны проводимости полупроводника n -типа, $I_T = I_{T_p}$ (рис. 106).

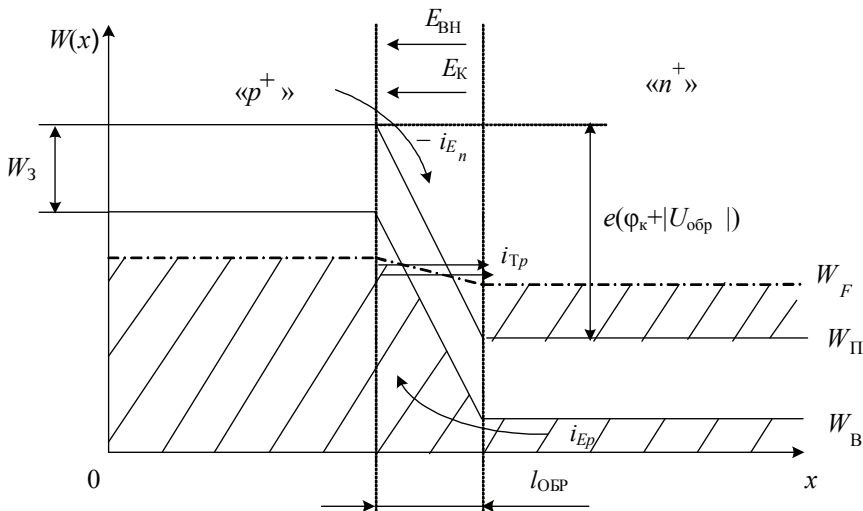


Рис. 106. Энергетическая диаграмма электронно-дырочного перехода с туннельным эффектом при $U = U_{\text{обр}}$

Обратный ток на участке OF (см. рис. 102) определяется по сумме $I_{\text{обр}} = I_E + I_{Tp}$.

Туннельная составляющая тока $I_{\text{обр}}$ с увеличением $|U_{\text{обр}}|$ возрастает по экспоненциальному закону, а величина дрейфовой составляющей обратного тока незначительна из-за малой концентрации неосновных носителей заряда в p - и n -областях электронно-дырочного перехода на основе «вырожденных» полупроводников. В точке F ВАХ (см. рис. 101) обратный ток равен току пика ($I_{\text{обр}} = I_{\Pi}$), и эта величина тока достигается при малых значениях обратного напряжения (десятки милливольт).

Вольт-амперная характеристика реального электронно-дырочного перехода с туннельным эффектом

В реальном электронно-дырочном переходе с туннельным эффектом в сильной степени изменяется прямая ветвь ВАХ (рис. 107). В основном изменяется значение напряжения впадины. Если для идеального p - n перехода с туннельным эффектом $U_{\text{вп}} \cong 2U_{\Pi}$, то для реального p - n перехода $U_{\text{вп}} \cong (5...7) U_{\Pi}$.

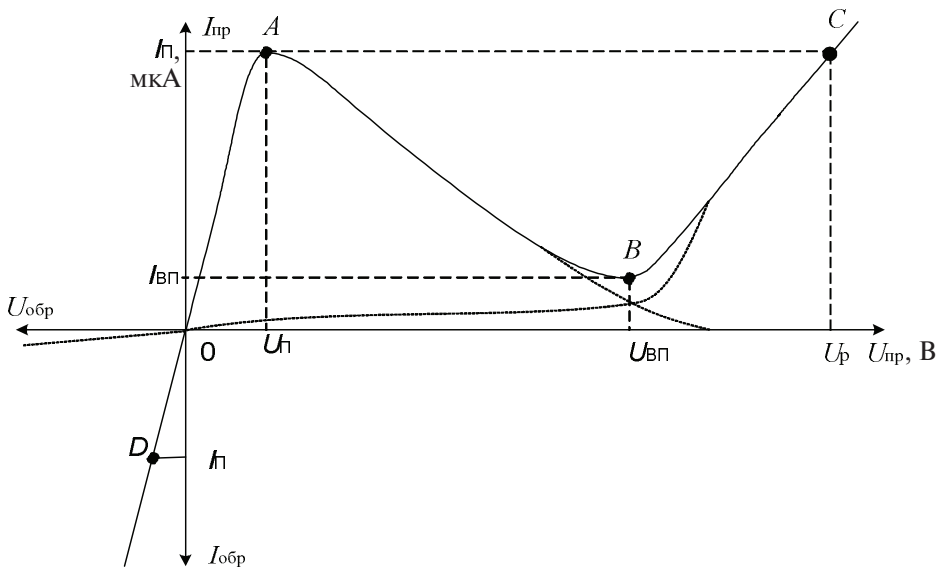


Рис. 107. Вольт-амперная характеристика реального p - n перехода с туннельным эффектом

Возрастание напряжения $U_{ВП}$ на рис. 107 вызвано наличием в p - n переходе избыточного туннельного тока. Электронно-дырочный переход, образованный «вырожденными» примесными полупроводниками, содержит в запрещенной зоне диаграммы локальные энергетические уровни примесных атомов, которые образуются из-за наличия неконтролируемых примесей в исходных полупроводниках.

Присутствие в запрещенной зоне локальных энергетических уровней таких примесных атомов приводит к появлению дополнительных туннельных переходов, а следовательно, избыточного туннельного тока (рис. 108).

При подаче прямого напряжения значением равным более двух значений напряжения пика ($U_{пр} > 2U_{П}$) на реальный p - n переход с туннельным эффектом имеют место туннельные переходы электронов с занятых энергетических уровней зоны проводимости полупроводника n -типа на локальные энергетические уровни примесных атомов в запрещенной зоне полупроводника p -типа, а с этих локальных энергетических уровней электроны переходят на свободные энергетические уровни валентной зоны полупроводника p -типа. Эти дополнительные туннельные переходы образуют избыточный туннельный ток p - n перехода, увеличивая тем самым участок туннельной составляющей прямого тока ВАХ электронно-дырочного перехода (см. рис. 107). В основном удлиняется «падающий участок» прямой ветви ВАХ (участок AB с отрицательным дифференциальным сопротивлением $r_{Диф АВ} < 0$).

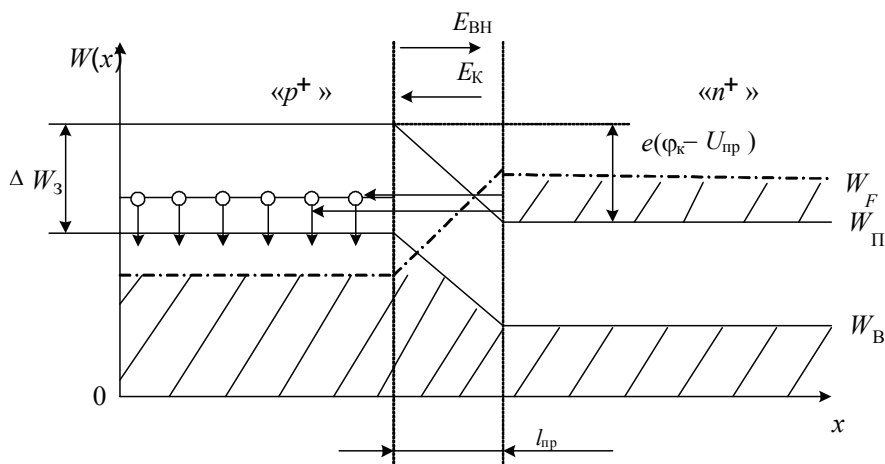


Рис. 108. Энергетическая диаграмма реального электронно-дырочного перехода с туннельным эффектом при $U_{пр} > 2U_{П}$

Дифференциальное сопротивление участка *AB* ВАХ (см. рис. 107) определяется по формуле

$$r_{\text{диф}_{AB}} = \frac{U_{\Pi} - U_{\text{ВП}}}{I_{\Pi} - I_{\text{ВП}}},$$

где U_{Π} — напряжение пика; $U_{\text{ВП}}$ — напряжение впадины; I_{Π} — ток пика; $I_{\text{ВП}}$ — ток впадины.

Точка *B* соответствует напряжению впадины ВАХ и образуется в месте пересечения диффузионной и туннельной составляющих прямого тока

$$I_{\text{ВП}} = I_{D(B)} + I_{\text{Тн}(B)} \mid U_{\text{пп}} = U_{\text{ВП}}.$$

Точка *C* ВАХ определяет напряжение раствора $U_{\text{р}}$. В точке *C* значение диффузионной составляющей прямого тока становится равной току пика ВАХ ($I_D = I_{\Pi}$). За счет смещения $U_{\text{ВП}}$ в сторону больших прямых напряжений напряжение раствора $U_{\text{р}}$ также возрастает по сравнению с его значением для ВАХ идеального *p-n* перехода с туннельным эффектом.

Электронно-дырочные переходы с туннельным эффектом изготавливаются с использованием полупроводниковых материалов на основе германия Ge и арсенида галлия GaAs, т. к. в этом случае получается лучшее отношение тока пика к току впадины ($I_{\Pi} / I_{\text{ВП}} > 5$), а *p-n* переходы с туннельным эффектом имеют параметры, приведенные в табл. 5. Электронно-дырочные переходы с туннельным эффектом на основе кремниевых полупроводников имеют малое отношение токов $I_{\Pi}/I_{\text{ВП}}$, поэтому на практике они не применяются.

Таблица 5

Параметры электронно-дырочного перехода с туннельным эффектом на основе германия и арсенида галлия

Параметры	Материал полупроводника <i>p-n</i> перехода	
	Ge	GaAs
U_{Π} , мВ	40–90	100–180
$U_{\text{ВП}}$, мВ	250–350	400–700
$U_{\text{р}}$, В	0,40–0,45	1,0–1,2
$\frac{I_{\Pi}}{I_{\text{ВП}}}$	5–20	8–40

Величина тока пика I_{Π} зависит от площади p - n перехода и степени легирования исходных «вырожденных» полупроводников. При увеличении концентрации примесей $N_{\Pi p}$ возрастает и ток пика I_{Π} .

Влияние температуры окружающей среды на ВАХ реального p-n перехода с туннельным эффектом

Колебания температуры окружающей среды приводят к изменению характеристик и параметров примесных полупроводников, а следовательно, и полупроводниковых структур на их основе. Из-за высокой степени легирования исходных примесных полупроводников электронно-дырочного перехода с туннельным эффектом, влияние температуры проявляется значительно слабее, чем в обычных p - n переходах. Особенно это заметно на участке вольт-амперной характеристики p - n перехода с туннельным эффектом, на котором ток определяется по туннельной составляющей общего тока p - n перехода

$$I = I_T + I_D + I_E,$$

где I — общий ток p - n перехода с туннельным эффектом; I_D — диффузионная составляющая тока; I_T — туннельная составляющая тока; I_E — дрейфовая составляющая тока. На ВАХ p - n перехода с туннельным эффектом, представленной на рис. 107, ток на участке $DOAB$ зависит в основном от туннельной составляющей. Поэтому на указанном участке ВАХ p - n перехода изменение температуры окружающей среды будет сказываться слабо, как это показано на рис. 109.

На участке OD ВАХ (рис. 109) обратный ток с ростом температуры окружающей среды увеличивается незначительно, т. к. при этом возрастает дрейфовая составляющая тока I_E , которая изменяется по экспоненциальному закону от температуры, а туннельная составляющая I_T изменяется очень слабо. Вес I_E в выражении для участка OD ВАХ очень мал. При увеличении температуры от 20 до 70 °С, экспоненциально возрастает дрейфовая составляющая тока I_E , уменьшается ширина обратносмещенного p - n перехода, растет напряженность электрического поля на p - n переходе ($E_{\text{обр}} = \frac{|U_{\text{обр}}|}{l_{\text{обр}}}$), несколько возрастает I_{Tp}

(туннельная составляющая тока обратносмещенного p - n перехода), это приводит к увеличению общего обратного тока $I_{\text{обр}} = I_{Tp} + I_E$, температурный коэффициент напряжения обратной ветви ВАХ p - n пере-

хода с туннельным эффектом имеет отрицательные значения

$$\text{ТКН}_{\text{обр}} = \frac{\Delta U_{\text{обр}}}{\Delta T} < 0.$$

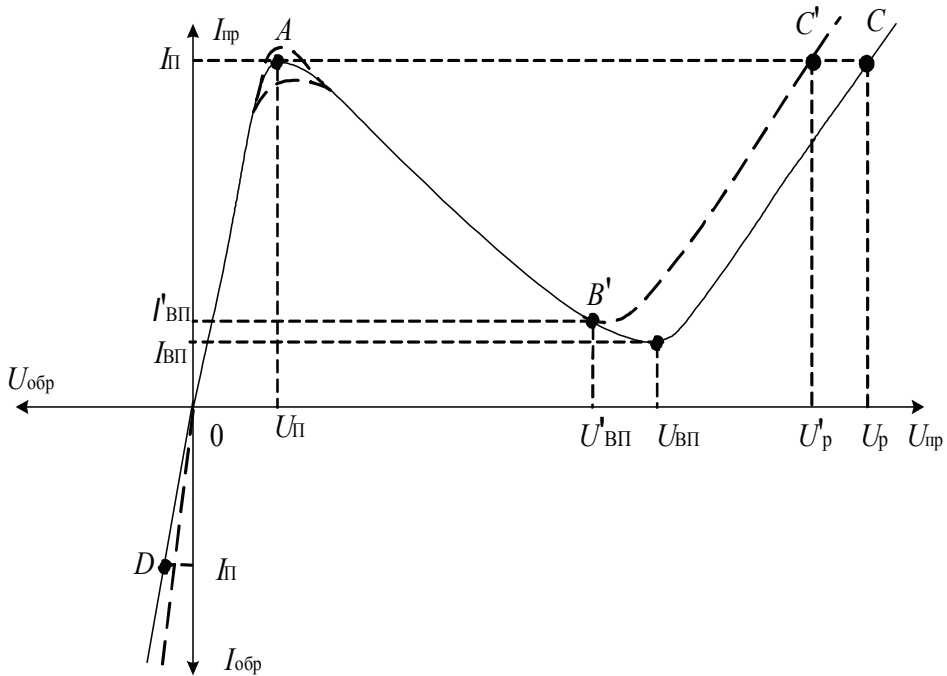


Рис. 109. ВАХ реального p - n перехода с туннельным эффектом при изменении температуры окружающей среды:
 — при $T_1 = +20\text{ }^{\circ}\text{C}$; - - - - - при $T_2 = +70\text{ }^{\circ}\text{C}$

При изменении температуры окружающей среды, обратная ветвь ВАХ располагается очень близко к первоначальной зависимости при $T_1 = +20\text{ }^{\circ}\text{C}$. Туннельная ветвь прямого тока ВАХ p - n перехода с туннельным эффектом (участок OAB зависимости на рис. 109) также слабо зависит от температуры окружающей среды. При возрастании температуры изменяется лишь ток пика I_{Π} . Он может как уменьшаться, так и увеличиваться. Это обусловлено следующими двумя факторами: во-первых, при возрастании температуры снижается значение контактной разности потенциалов ϕ_k p - n перехода, вследствие чего уменьшается равновесная ширина p - n перехода l_0 и l_{Π} , что приводит к росту напряженности электрического поля E и туннельной составляющей тока; во-вторых, при увеличении температуры происходит смещение энер-

гетического уровня Ферми ближе к середине запрещенной зоны (это следует из теории примесных полупроводников) в полупроводниках p^+ - и n^+ -типов, а это приводит к уменьшению перекрытия зон в электронно-дырочном переходе, содержащих энергетические уровни, заполненные электронами. При этом количество туннельных переходов снижается, а также уменьшаются значения токов I_{Tn} и I_{Π} .

На диффузионном участке (участок BC) прямой ветви ВАХ электронно-дырочного перехода с туннельным эффектом влияние температуры проявляется так же, как и для обычного p - n перехода. С возрастанием температуры снижается значение контактной разности потенциалов ϕ_k , растет значение диффузионной составляющей тока I_D и возрастает величина прямого тока $I_{пр}$ p - n перехода. Поэтому диффузионный участок BC ВАХ p - n перехода с увеличением температуры ($T_2 = +70^\circ\text{C}$) смещается влево. В этом случае увеличивается значение тока впадины и он становится равным $I'_{вп}$, уменьшается величина напряжения впадины до значения $U'_{вп}$, снижается напряжение раствора до U'_p . Температурный коэффициент напряжения участка BC прямой ветви ВАХ меньше нуля и определяется из соотношения

$$\text{ТКН}_{пр(BC)} = \frac{\Delta U_{пр}}{\Delta T} < 0.$$

Параметры электронно-дырочного перехода с туннельным эффектом

Параметры электронно-дырочных переходов с туннельным эффектом имеют некоторую специфику по сравнению с параметрами, применяемыми для характеристики обычных p - n переходов. В качестве параметров используют напряжения и токи, определяющие характерные точки вольт-амперной характеристики (см. рис. 107). К ним относятся I_{Π} , U_{Π} , $I_{вп}$, $U_{вп}$, U_p , $\frac{I_{\Pi}}{I_{вп}}$, $U_{обр доп}$.

Пиковый ток I_{Π} — это прямой ток в точке максимума прямой ветви вольт-амперной характеристики (точка A на рис. 107). Величина тока пика зависит, во-первых, от степени легирования областей, т. к. это определяет глубину залегания энергетических уровней Ферми в разрешенных зонах исходных полупроводников n^+ - и p^+ -типов; во-вторых,

пропорциональна площади электронно-дырочного перехода. При увеличении концентрации примесей уменьшается равновесная ширина p - n перехода l_0 , а контактная разность потенциалов ϕ_k растет, что приводит к увеличению напряженности электрического поля в переходе и значительному возрастанию тока пика. Величина тока пика для различных электронно-дырочных переходов с туннельным эффектом колеблется от единиц миллиампер до нескольких ампер.

Напряжение пика U_{Π} — это прямое напряжение, соответствующее пиковому току I_{Π} (точка A ВАХ на рис. 107). Оно зависит от степени легирования и составляет обычно 60–90 мВ для германиевых и 100–180 мВ арсенид-галлиевых p - n переходов.

Ток впадины $I_{ВП}$ — это прямой ток в точке минимума ВАХ. Ток впадины зависит от материала полупроводника (энергии ширины запрещенной зоны), из которого изготовлен электронно-дырочный переход, и температуры окружающей среды.

Напряжение впадины $U_{ВП}$ — это прямое напряжение, соответствующее току впадины $I_{ВП}$ (точка B ВАХ на рис. 107). Напряжение впадины тем больше, чем выше энергия ширины запрещенной зоны полупроводника. Для германиевых p - n переходов с туннельным эффектом $U_{ВП} = 250...300$ мВ, а для p - n переходов, выполненных на основе арсенида галлия, $U_{ВП} = 500...700$ мВ.

Напряжение раствора U_p — это прямое напряжение, которое больше напряжения впадины и соответствует прямому току равному току пика. Напряжение раствора зависит от материала полупроводника, из которого изготовлен p - n переход. Обычно $U_{p\text{ Ge}} = 400...450$ мВ; а $U_{p\text{ GaAs}} = 1000...1200$ мВ.

Отношение пикового тока к току впадины $I_{\Pi}/I_{ВП}$ характеризует крутизну падающего участка прямой ветви ВАХ. Это отношение зависит от материала полупроводника, на основе которого выполнен p - n переход с туннельным эффектом, и составляет для германиевых переходов 5–10, а арсенид-галлиевых 8–40.

Максимальнодопустимое обратное напряжение $U_{\text{обр доп}}$ соответствует значению обратного тока $I_{\text{обр}(D)} = -I_{\Pi}$ (точка D ВАХ на рис. 107).

Кроме параметров, определяемых значениями токов и напряжений ВАХ электронно-дырочного перехода с туннельным эффектом, определяются сопротивления постоянному и переменному токам для прямой и обратной ветвей вольт-амперной характеристики (см. рис. 108).

Сопротивление постоянному току R_0 для прямой ветви ВАХ обычно определяется для точки C

$$R_{0(C)} = \frac{U_P}{I_P}.$$

Для обратной ветви ВАХ R_0 определяется для точки D

$$R_{0(D)} = \left| \frac{U_{\text{обр_доп}}}{I_P} \right|.$$

При определении сопротивления переменному току $r_{\text{диф}}$ для прямой ветви ВАХ выделяются три характерных участка OA , AB , BC . Для участка OA

$$r_{\text{диф } (OA)} = \frac{\Delta U_{\text{пр1}}}{\Delta I_{\text{пр1}}} > 0;$$

для участка AB (падающий участок прямой ветви ВАХ)

$$r_{\text{диф } (AB)} = \frac{U_P - U_{\text{ВП}}}{I_P - I_{\text{ВП}}} < 0;$$

для участка BC (участок диффузионной составляющей прямого тока ВАХ *p-n* перехода)

$$r_{\text{диф } (BC)} = \frac{\Delta U_{\text{пр2}}}{\Delta I_{\text{пр2}}} > 0.$$

Для обратной ветви ВАХ (участок OD ВАХ) дифференциальное сопротивление находится из соотношения

$$r_{\text{диф } (OD)} = \frac{\Delta U_{\text{обр}}}{\Delta I_{\text{обр}}} > 0.$$

Задачи к практическому занятию № 3

- 1) Имеется германиевый *p-n* переход с концентрацией примесей $N_D = 10^3 N_A$, причем на каждые 10^8 атомов германия приходится один атом акцепторной примеси. Определить контактную разность потенциалов при температуре $T = 300$ К.

- 2) В структуре с германиевым p - n переходом удельная проводимость p -области $\gamma_p = 10^4$ См/м и удельная проводимость n -области $\gamma_n = 10^2$ См/м. Вычислить контактную разность потенциалов в переходе при температуре $T = 300$ К.
- 3) Используя данные и результаты предыдущей задачи, найти: а) плотность обратного тока насыщения, а также отношение дырочной составляющей обратного тока насыщения к электронной, если диффузионная длина электронов и дырок $L_n = L_p = 10^{-3}$ м; б) напряжение, при котором плотность прямого тока $j = 10^5$ А/м².
- 4) Германиевый p - n переход имеет обратный ток насыщения 1 мкА; кремниевый переход таких же размеров — обратный ток насыщения 10^{-8} А. Вычислить и сравнить прямые напряжения $U_{пр}$ на переходах при $T = 293$ К и токе 100 мА.
- 5) Концентрация доноров и акцепторов в n - и p -областях резкого p - n перехода соответственно равна $5 \cdot 10^{16}$ см⁻³ и 10^{17} см⁻³. Определить контактную разность потенциалов и плотность обратного тока насыщения, полагая, что при комнатной температуре коэффициенты диффузии для неосновных электронов и дырок составляют 100 и 50 см²/с соответственно, а диффузионная длина $L_n = L_p = 0,8$ см. Собственную концентрацию носителей заряда считать равной 10^{13} см⁻³.

Задания: решение задач № 1–4 раздела «Электронно-дырочный переход» (приложение «Домашние задания»).

Задачи к ПРАКТИЧЕСКОМУ ЗАНЯТИЮ № 4

- 1) Германиевый полупроводниковый диод, имеющий обратный ток насыщения $I_0 = 25$ мкА, работает при прямом напряжении равном 0,1 В и $T = 300$ К. Определить: а) сопротивление диода постоянному току R_0 ; б) дифференциальное сопротивление $r_{диф}$.
- 2) Определить, во сколько раз увеличивается обратный ток насыщения сплавного p - n перехода диода, если температура увеличивается: а) от 20 до 80 °С для германиевого перехода; б) от 20 до 150 °С для кремниевого перехода.

- 3) При прямом напряжении 0,1 В на *p-n* переходе, через него проходит определенный ток. Каким должно быть прямое напряжение, чтобы ток увеличился в два раза? Расчет провести для комнатной температуры.
- 4) Барьерная емкость резкого *p-n* перехода равна 200 пФ при обратном напряжении 2 В. Какое требуется обратное напряжение, чтобы она уменьшилась до 50 пФ, если контактная разность потенциалов $\phi_k = 0,82$ В?
- 5) Во сколько раз изменится барьерная емкость резкого *p-n* перехода при увеличении обратного напряжения от 20 до 80 В?

Задание: определяются индивидуальные исходные данные для выполнения расчетно-графической работы «Расчет *p-n* перехода».

9. ГЕТЕРОПЕРЕХОДЫ

9.1. ОБЩИЕ СВЕДЕНИЯ О ГЕТЕРОПЕРЕХОДАХ

Гетеропереход — это переходный слой с существующим в нем диффузионным электрическим полем между двумя различными по химическому составу полупроводниками.

На границе раздела обычно изменяется ширина запрещенной зоны, подвижность носителей заряда, их эффективная масса и другие характеристики полупроводников. Поскольку гетеропереходы изготавливаются из двух различных материалов, их степень совершенства зависит от кристаллической структуры материалов, близости постоянных решетки, сорта и концентрации легирующих примесей, непрерывности и полярности связей, многих особенностей технологии изготовления. Надежная работа гетероперехода в качестве элемента прибора требует совпадения механических и тепловых констант материалов, в первую очередь коэффициента линейного расширения в широком интервале температур. Близость постоянных решетки является первым и важнейшим критерием совместимости материалов пары. Из-за этих факторов количество материалов для изготовления гетеропереходов ограничено. Для данных целей подходят, например, соединения типа $A^{II}B^{VI}$, $A^{III}B^V$. Наиболее распространенные пары веществ для гетеропереходов — это Ge-GaAs, GaAs-GaP, GaAs-IAs, GaAs-AlGaAs.

Различают *n-p* и *p-n* гетеропереходы (на первое место ставится буква, обозначающая тип электропроводности полупроводника с более узкой запрещенной зоной). На основе гетеропереходов возможно также создание структур *n-n* и *p-p*.

9.2. ЭНЕРГЕТИЧЕСКАЯ ДИАГРАММА ГЕТЕРОПЕРЕХОДА

Отличия свойств гетеропереходов от переходов в монокристаллах (гомопереходов) видны при получении их энергетических диаграмм. Рассмотрим энергетическую диаграмму идеального гетероперехода (рис. 110).

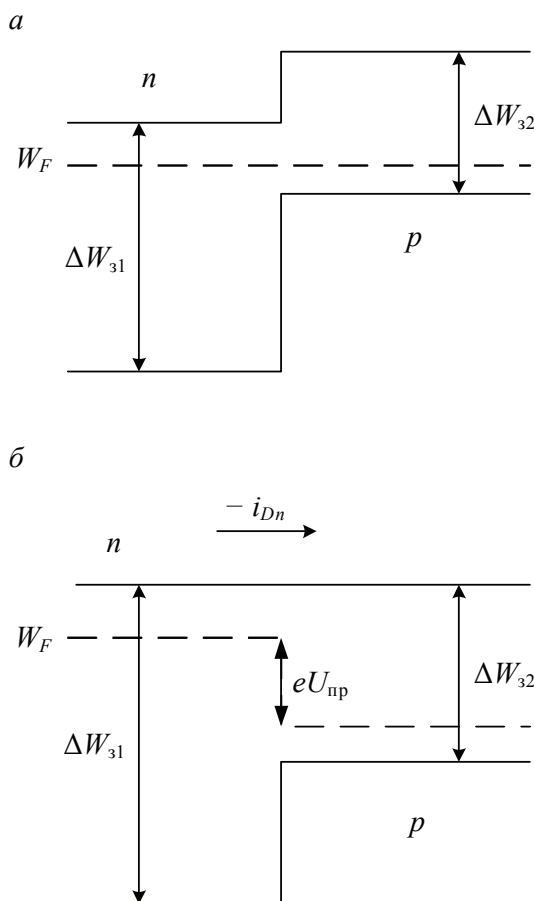


Рис. 110. Энергетическая диаграмма идеального гетероперехода:
 а — переход в равновесном состоянии; б — прямосмещенный переход

Каждый из полупроводников, образующих гетеропереход, может иметь различный тип электропроводимости. Поэтому для каждой пары полупроводников можно осуществить четыре типа гетеропереходов:

$p_1-p_2, p_1-n_2, n_1-p_2, n_1-n_2$. Но все они образуются по одному принципу. При образовании гетеропереходов, из-за разных работ выхода электронов из разных полупроводников, происходит перераспределение носителей заряда в приконтактной области и выравнивание уровней Ферми в результате установления термодинамического равновесия. Все остальные энергетические уровни и зоны должны соответственно изогнуться, т. е. в гетеропереходе возникает диффузионное электрическое поле и контактная разность потенциалов. При этом энергетический уровень потолка верхней свободной зоны должен быть непрерывным.

Обычно энергетический уровень потолка верхней свободной зоны является энергетическим уровнем потолка зоны проводимости, т. к. свободные энергетические зоны перекрывают друг друга. Разрыв дна зоны проводимости определяется различием энергии сродства к электрону двух контактирующих полупроводников (энергия сродства к электрону есть разница энергии потолка верхней свободной зоны и дна зоны проводимости). Разрыв потолка валентной зоны зависит как от разницы энергии сродства, так и от различия ширины запрещенных зон контактирующих полупроводников.

В результате разрывов дна зоны проводимости и потолка валентной зоны, высота потенциальных барьеров для электронов и дырок в гетеропереходе оказывается различной. Это и является особенностью гетеропереходов, обуславливающей их специфические свойства.

Рассмотрим p - n гетеропереход, у которого ширина запрещенной зоны n -полупроводника больше, чем у p -полупроводника (рис. 110). Важным параметром, характеризующим пару полупроводников, является разность энергии дна зоны проводимости

$$\Delta W_{\Pi} = W_{\Pi 1} - W_{\Pi 2}.$$

Разность энергии потолка валентной зоны выражается через ΔW_{Π} :

$$\Delta W_B = W_{B2} - W_{B1} = \Delta W_{31} - \Delta W_{32} - \Delta W_{\Pi}.$$

Для энергетической диаграммы реального гетероперехода (рис. 111) характерны разрывы границ зон равные ΔW_{Π} и ΔW_B на металлургической границе — x_0 . Изгибы границ зон вблизи x_0 связаны с образованием обедненных слоев толщиной l_n и l_p , содержащих объемные заряды ионов доноров и акцепторов. Значение изгибов ($q\varphi_{01}$, $q\varphi_{02}$) равно внутренней разности потенциалов, образующейся в обедненных слоях,

$$\varphi_0 = \varphi_{01} + \varphi_{02} = (W_{F1} - W_{F2})/q$$

представляет собой контактную разность потенциалов. Как и для гомоперехода, она может быть выражена через уровни Ферми в изолированных полупроводниках.

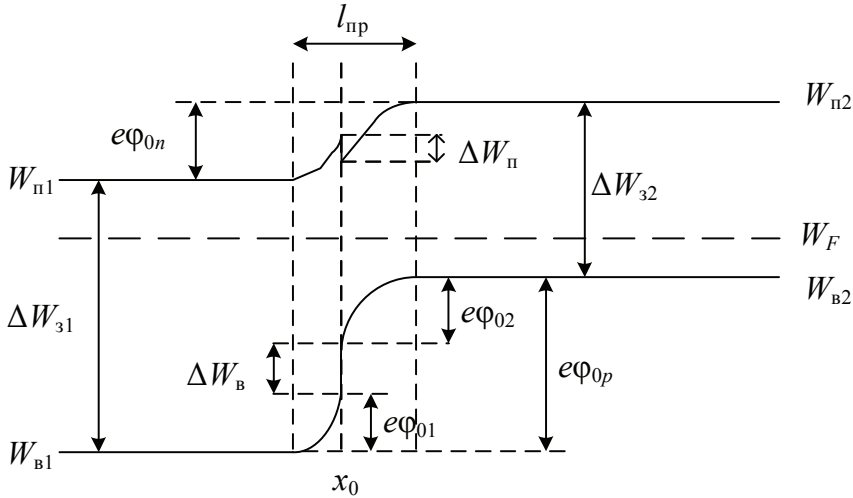


Рис. 111. Энергетическая диаграмма реального гетероперехода в равновесном состоянии

Однако для гетероперехода ϕ_0 не совпадает с высотой потенциального барьера. Высота потенциального барьера для электронов, движущихся из n - в p -область, будет

$$\phi_{0n} = \phi_0 - \Delta W_{\pi}/q,$$

для дырок, движущихся из p - в n -область, —

$$\phi_{0p} = \phi_0 + \Delta W_{\text{в}}/q.$$

Таким образом, для электронов барьер ниже, чем для дырок, на величину

$$\phi_{0p} - \phi_{0n} = (\Delta W_{31} - \Delta W_{32})/q.$$

Если вблизи границы раздела двух полупроводников, образующих гетеропереход, возникают обедненные основными носителями слои с повышенным сопротивлением, то основная часть внешнего напряжения, приложенного к структуре с гетеропереходом, будет падать на обедненных слоях. Высота потенциального барьера будет изменяться: уменьшаться при полярности внешнего напряжения, противоположной полярности контактной разности потенциалов, и увеличивать-

ся при совпадении их полярностей. Таким образом, гетеропереходы могут обладать эффектом выпрямления.

Из-за различия по высоте потенциальных барьеров для электронов и для дырок, прямой ток гетероперехода связан в основном с движением носителей заряда только одного знака. Поэтому при приложении прямого напряжения к переходу будет преобладать инжекция электронов, даже если p -область имеет такую же (а иногда и большую) концентрацию основных носителей, что и n -область. Таким образом, можно получить коэффициент инжекции γ близкий к 1 (т. е. одностороннюю инжекцию неосновных носителей заряда из широкозонного в узкозонный полупроводник)

$$\gamma = \frac{1}{1 + \frac{\sigma_B}{\sigma_9}},$$

где σ_B и σ_9 — удельные проводимости базы и эмиттера соответственно.

Это явление называется суперинжекцией.

В гетеропереходах, образованных полупроводниками с одним типом электропроводности p_1 - p_2 , n_1 - n_2 , выпрямление происходит без инжекции неосновных носителей заряда.

Зонная модель идеального гетероперехода может быть использована для реальных гетеропереходов с определенным приближением, т. к. почти всегда на границе раздела гетероперехода присутствуют заряженные поверхностные состояния (уровень Ферми оказывается фиксированным, вместо ровного хода для одной из зон имеет место барьер типа Шоттки). Такие состояния, несмотря на развитие эпитаксиальной технологии, неизбежно возникают при образовании гетеропереходов вследствие появления краевых дислокаций из-за несоответствия кристаллических решеток.

9.3. ПРЯМОСМЕЩЕННЫЙ ГЕТЕРОПЕРЕХОД

Приложение прямого смещения к гетеропереходу, как упоминалось выше, будет обеспечивать одностороннюю эффективную инжекцию из материала с большей шириной запрещенной зоны. Так, в рассматриваемом гетеропереходе (n -часть широкозонная) (рис. 112) наличие разрыва будет препятствовать инжекции дырок в широкозонный полупроводник.

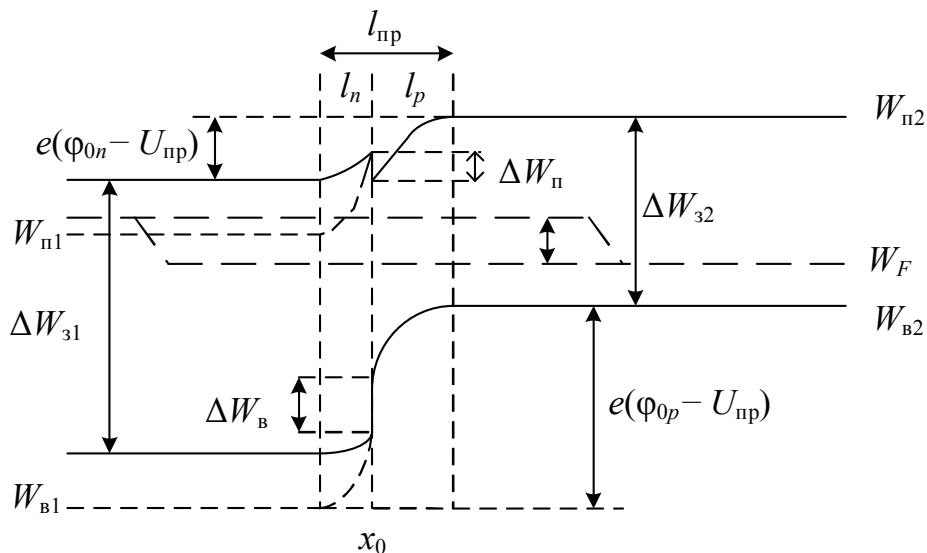


Рис. 112. Прямосмещенный гетеропереход

Наличие разрыва в зоне проводимости уменьшает барьер для инжектируемых электронов. Через переход течет ток, и система становится неравновесной. При этом стационарный уровень Ферми не существует, а концентрацию носителей заряда принято описывать с помощью квазиуровней Ферми. Благодаря скачку дна зоны проводимости, электронный квазиуровень проникает в p -зону и оканчивается там выше, чем при условии равновесия. Так концентрация неравновесных электронов оказывается даже выше, чем в n -полупроводнике. Такой рост концентрации неравновесных электронов в p -полупроводнике при протекании тока называют суперинжекцией.

При прямом смещении ширина перехода уменьшается так же, как и при монокристаллическом переходе.

9.4. ПОЛУЧЕНИЕ ГЕТЕРОПЕРЕХОДОВ

Широкое практическое использование нашли эффективно инжектирующие гетеропереходы, близкие по свойствам к идеальной модели, на полупроводниковых соединениях $A^{\text{III}}B^{\text{V}}$ и их твердые растворы на основе арсенидов, фосфидов и антимонидов галлия и алюминия. Благодаря совпадению ковалентных радиусов галлия и алюминия гетеропе-

переходов, изменение их химического состава и ширины запрещенной зоны происходит без изменения периода кристаллической решетки.

На основе твердых растворов соединений $A^{III}B^V$ можно создать практически любое число гетеропереходов с совпадающими постоянными решетки, однако их получение требует развития очень сложных методов кристаллизации.

Среди весьма большого числа различных методов эпитаксиального роста полупроводниковых кристаллов наибольшее распространение получили методы:

- 1) химических транспортных реакций;
- 2) испарения в вакууме;
- 3) выращивания из раствора-расплава;
- 4) поверхностного оплавления.

В первом из этих методов обычно галогенами транспортируются атомы выращиваемого вещества в зону кристаллизации. Таким способом были получены первые монокристаллические гетеропереходы арсенид галлия — германий. Методы испарения в вакууме и поверхностного оплавления пока не получили широкого распространения из-за трудностей получения этими способами совершенной границы раздела в гетеропереходах.

Одним из самых простых и эффективных оказался метод эпитаксиального выращивания из раствора-расплава, получивший также название метода жидкостной эпитаксии.

В кварцевую трубу в токе водорода помещают в графитовой кассете кристалл-подложку, на которую наливают металл-растворитель, хорошо смачивающий поверхность подложки. В металл-растворитель добавляют вещество, кристаллизацию которого необходимо провести на подложке. Особенно удобен этот метод для получения гетеропереходов между полупроводниковыми соединениями, в состав которых входит сам металл-растворитель. Методом жидкостной эпитаксии были получены гетеропереходы в системе $AlAs-GaAs$ и многослойные (четыре и более слоев) гетероструктуры на их основе.

9.5. ЭЛЕКТРИЧЕСКИЕ СВОЙСТВА ГЕТЕРОПЕРЕХОДОВ

Рассмотрим вольт-амперные характеристики анизотипных гетеропереходов $p-Si/n-CdTe$. На рис. 113 приведена вольт-амперная характеристика гетероструктуры $pSi/n-CdTe$ с $pSi = 5 \cdot 10^{14} \text{ см}^{-3}$,

$n\text{CdTe} = 1 \cdot 10^{15} \text{ см}^{-3}$, которая является типичной для серии гетероструктур, полученных в оптимальных условиях конденсации $T_{\text{конд}} = 350 \dots 380 \text{ }^\circ\text{C}$ из газодинамического потока.

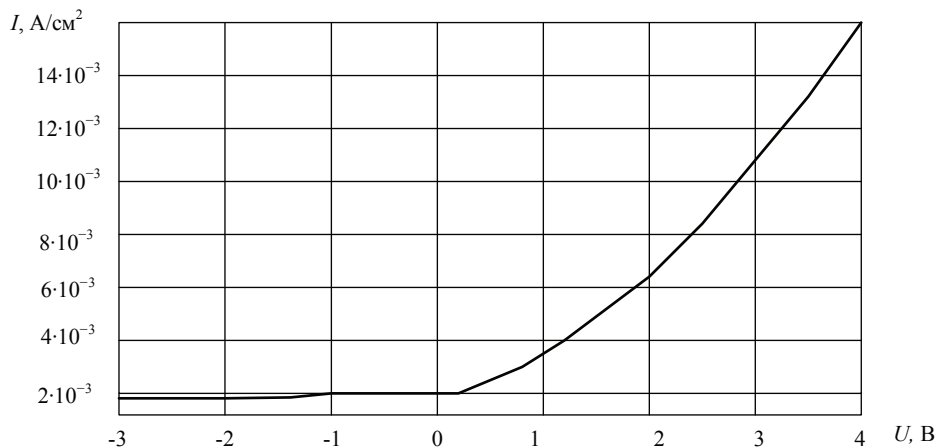


Рис. 113. Вольт-амперная характеристика анизотипного гетероперехода

Гетеропереход обладает выпрямляющими свойствами с высокими значениями коэффициента выпрямления. При малых напряжениях прямого смещения (обычно до нескольких десятых вольта), на начальном участке ВАХ удовлетворительно описывается соотношением

$$I = I_0 \left[\exp \left(\frac{eU}{\beta kT} - 1 \right) \right],$$

соответствующим термоэмиссионным токам, где β — коэффициент неидеальности. В классической теории, учитывающей рекомбинацию в области пространственного заряда, $\beta = 1 \dots 2$. Для рассматриваемого гетероперехода $\beta = 2,3$, что указывает на его неидеальность, которая проявляется, вероятнее всего, с влиянием на токопротекание граничных состояний. Для других гетероструктур, полученных в оптимальных условиях, $\beta = 2,0 \dots 4,44$.

В области больших положительных смещений (более 2 В), зависимость тока от напряжения описывается степенным законом $I \sim U^m$ с показателем $m = 1,5 \dots 2,0$. Такие значения обычно связываются с инжекционными токами и определяются сопротивлением и наличием дефектных центров в базовой области CdTe.

В широком диапазоне напряжений обратная вольт-амперная характеристика описывается соотношением

$$I \sim a \exp \frac{b}{\sqrt{\Phi_0 - eU}},$$

где a и b — константы, не зависящие от напряжения и температуры. Такое соотношение указывает на то, что доминирующим механизмом протекания тока является туннелирование.

Рассмотрим гетеропереходы, образованные соединениями $A^{II}B^{VI}$, несоответствие решеток между которыми больше 7 % (это гетеропереходы с большим несоответствием решеток). К таким гетеропереходам относятся ZnSe-ZnTe и ZnSe-CdTe. Каждый из этих гетеропереходов изготавливался различными методами, однако их ВАХ имели много общего. При получении гетеропереходов ZnSe-ZnTe вакуумной эпитаксией, наиболее подробное изучение электрических свойств проводилось на гетеропереходах, полученных при сравнительно высоких температурах подложки, при которых образуются совершенные по структуре слои. На полученных ВАХ в пределах пяти порядков изменения тока наблюдается экспоненциальная зависимость тока от напряжения. При небольших положительных смещениях, логарифм тока пропорционален температуре образца.

Гетеропереход ZnSe-CdTe получается двух видов в зависимости от того, какой из материалов являлся подложкой. При выращивании теллурида кадмия на подложке из селенида цинка, структура эпитаксиальных слоев оказывается более совершенной, и вследствие этого изменяется и механизм токопрохождения. ВАХ таких образцов резко несимметричны, коэффициент выпрямления при 300 К и $U_{пр} = 2$ В достигает значений 10^{16} .

9.6. ПРИМЕНЕНИЕ ГЕТЕРОПЕРЕХОДОВ В ОПТОЭЛЕКТРОНИКЕ

Кроме свойств выпрямления гетеропереходов, для полупроводниковых приборов интересными и полезными оказываются различия спектров поглощения и показателей преломления образующих гетеропереход полупроводников.

В полупроводниковом инжекционном лазере с гетеропереходами просто создать инверсную населенность (рис. 114). Базовую область такой структуры делают из полупроводника меньшей ширины запрещенной зоны и большей диэлектрической проницаемостью, чем у эмиттер-

ных областей. Инжектированные в базу носители заряда оказываются в потенциальных ямах. Различие показателей преломления (диэлектрических проницаемостей) базовой и эмиттерной областей приводит к полному внутреннему отражению квантов света на гетеропереходах, т. е. область базы является по существу световодом. Все это обеспечивает значительно меньшие пороговые плотности токов и большие эффективность или КПД инжекционных лазеров с гетеропереходами.

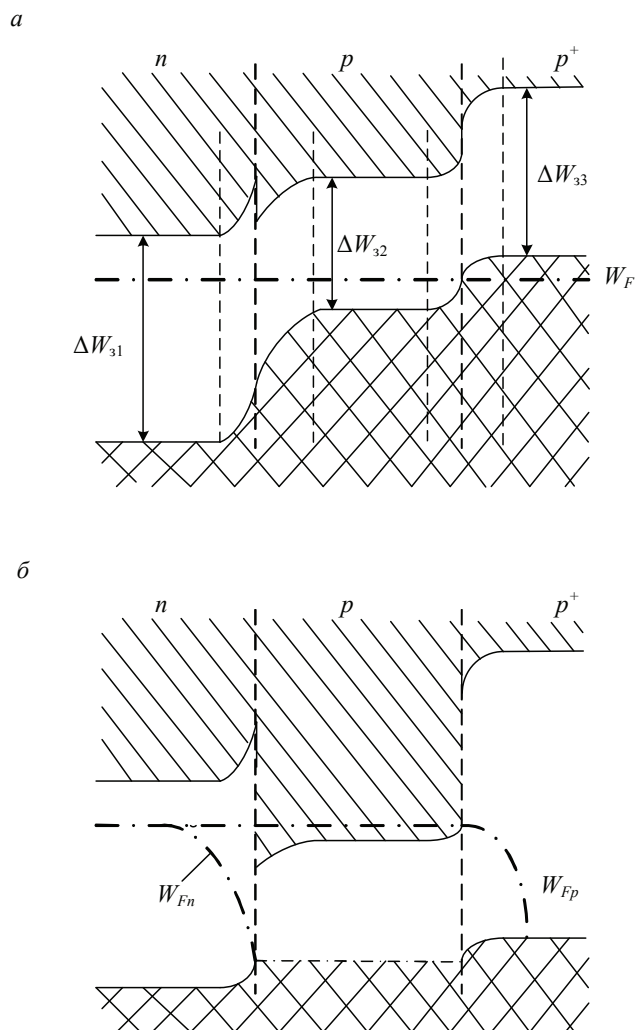


Рис. 114. Энергетическая диаграмма полупроводникового инжекционного лазера с гетеропереходами без проложенного напряжения (*a*) и при прямом напряжении (*б*)

Инжекционный лазер представляет собой полупроводниковый двухэлектродный прибор с p - n переходом (поэтому часто как равноправный используется термин «лазерный диод»), в котором генерация когерентного излучения связана с инъекцией носителей заряда при протекании прямого тока через p - n переход.

Рассмотренные теоретические положения определяют пути совершенствования простейшей структуры лазера. Исследованы и реализованы варианты расположения слоев по толщине кристалла. В гомогенном полупроводнике p - n переход, как средство электронного ограничения, весьма несовершенен: при высоких уровнях накачки происходит бесполезная инжекция электронов влево (из-за падения коэффициента инжекции), ограничение справа достигается лишь естественным убыванием концентрации введенных дырок по экспоненциальному закону. Границы, определяющие электронную и оптическую толщины активной области, непостоянны. Все эти несовершенства, проявляющиеся, в конечном счете, в высоком значении плотности порогового тока, предопределили бесперспективность лазеров на однородных полупроводниках.

Широкое промышленное распространение получили только гетеролазеры, особенностями которых являются односторонняя инжекция, четко выраженный волноводный эффект, возможность суперинжекции.

В односторонней гетероструктуре (ОГС) электронное ограничение слева идеально, а справа такое же, как и в лазере на гомогенном полупроводнике; преимущество ОГС перед другими гетероструктурами состоит в простоте технологии.

Существует еще двойная (двусторонняя) гетероструктура (ДГС), в которой сверхтонкая активная область «зажата» между двумя гетерограницами — именно она позволяет получать малые пороговые плотности тока и значительные выходные мощности. Четырех- и пятислойная структура, являющаяся усовершенствованной ДГС, позволяет иметь оптимальную толщину. Отметим, что технологические соображения требуют создания ряда переходных слоев, поэтому реальные лазерные структуры значительно сложнее, чем физические модели.

Инжекционные лазеры имеют ряд достоинств, выделяющих их среди излучателей и предопределяющих доминирующую роль в оптоэлектронике:

- 1) микроминиатюрность — теоретическая минимальная длина резонатора близка к 10 мкм, а площадь его поперечного сечения

1–2 мкм (объем активной области может достигать 10–12 см³). Это возможно потому, что в полупроводниковых лазерах индуцированные переходы связаны не с отдельными дискретными уровнями, а с переходами зона — зона, поэтому и усиление в них наибольшее;

- 2) высокий КПД преобразования энергии накачки в излучение, приближающийся у лучших образцов к теоретическому пределу. Это обусловлено тем, что лишь при инжекционной накачке удастся исключить нежелательные потери — вся энергия электрического тока переходит в энергию возбужденных электронов;
- 3) удобство управления — низкие напряжение и ток возбуждения, совместимые с интегральными микросхемами; возможность изменения мощности излучения без применения внешних модуляторов; работа как в непрерывном, так и в импульсном режимах с обеспечением при этом очень высокой скорости переключения (в пикосекундном диапазоне);
- 4) возможность генерации требуемой спектральной линии, обеспечиваемую выбором или синтезом полупроводника с необходимой шириной запрещенной зоны;
- 5) пригодность для массового производства, низкую стоимость, долговечность;
- 6) совместимость с основным элементом микроэлектроники — транзистором (по типу используемых материалов и технологии обработки). Это открывает принципиальную возможность создания интегрированных лазерных излучателей.

Инжекционным лазерам присущи и определенные недостатки, к принципиальным можно отнести следующие:

- 1) невысокую когерентность излучения (в сравнении, например, с газовыми лазерами) — значительную ширину спектральной линии, большую угловую расходимость, эллиптический астигматизм;
- 2) относительно малую генерируемую мощность (некоторым оптоэлектронным устройствам, например голографическим ЗУ, требуются лазеры большой мощности);
- 3) существенность таких негативных явлений, как временная деградация (в особенности для коротковолновых лазеров), резкое уменьшение мощности излучения при повышении температуры и воздействии радиации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Федотов, Я. А. Основы физики полупроводниковых приборов / Я. А. Федотов. — Москва : Сов. радио, 1969. — 542 с.
2. Пасынков, В. В. Полупроводниковые приборы : учебник для вузов / В. В. Пасынков, Л. К. Чиркин. — 5-е изд., испр. — Санкт-Петербург : Лань, 2001. — 480 с. — ISBN 5-8114-0368-2.
3. Электронные приборы : учебник / В. Н. Дулин, Н. А. Аваев, В. П. Демин [и др.] ; под ред. Г. Г. Шишкина. — 4-е изд., перераб. и доп. — Москва : Энергоатомиздат, 1989. — 496 с. — ISBN 5-283-01472-X.
4. Булычев, А. Л. Электронные приборы : учебник / А. Л. Булычев, П. М. Лямин, Е. С. Тулинов. — Москва : Лайт Лтд., 2001. — 416 с. — ISBN 5-89818-048-6.
5. Батушев, В. А. Электронные приборы : учебник / В. А. Батушев. — 2-е изд., перераб. и доп. — Москва : Высшая школа, 1980. — 383 с.
6. Елфимов, В. И. Основы теории p - n перехода : учебное пособие / В. И. Елфимов, Н. С. Устыленко. — Екатеринбург : Издательство УМЦ УПИ, 2000. — 55 с.
7. Бондаренко, И. Б. Электрорадиоэлементы. Часть 1. Резисторы : учебное пособие / И. Б. Бондаренко. — Санкт-Петербург : СПб НИУ ИТМО, 2012. — 108 с.
8. Пасынков, В. В. Материалы электронной техники : учебник для вузов / В. В. Пасынков, В. С. Сорокин. — Санкт-Петербург : Лань, 2001. — 368 с. — ISBN 5-8114-0409-3.
9. Антипов, Б. Л. Материалы электронной техники. Задачи и вопросы : учебное пособие для вузов / Б. Л. Антипов, В. С. Сорокин, В. А. Терехов ; под ред. В. А. Терехова. — 2-е изд. — Санкт-Петербург : Лань, 2001. — 208 с. — ISBN 5-8114-0410-7.
10. Резисторы : справочник / В. В. Дубровский, Д. М. Иванов, Н. Я. Пратусевич [и др.] ; под ред. И. И. Четверткова и В. М. Те-

- рехова. — 2-е изд., перераб. и доп. — Москва : Радио и связь, 1991. — 528 с. — ISBN 5-256-00374-7.
11. Дурнаков, А. А. Электроника : учебно-методическое пособие / А. А. Дурнаков, В. И. Елфимов. — Екатеринбург : Изд-во Урал. ун-та, 2016. — 160 с. — ISBN 978-5-7996-1787-5.
 12. Парфенова, Е. Л. Физические основы микро- и нанoeлектроники : учебное пособие для студентов вузов, обучающихся по специальности 200101 «Приборостроение» / Е. Л. Парфенова, Л. А. Терентьева, М. Г. Хусаинов. — Ростов-на-Дону : Феникс, 2012. — 234 с. — ISBN 978-5-222-19617-5.
 13. Дробот, П. Н. Нанoeлектроника : учебное пособие / П. Н. Дробот. — Томск : ТУСУР, 2016. — 286 с.

ПРИЛОЖЕНИЕ

ДОМАШНИЕ ЗАДАНИЯ

Цель домашних заданий — активизировать самостоятельную учебную работу студентов, сформировать глубокое понимание физической сущности процессов и явлений в материалах электронной техники, закрепить основные формулы и определения, а также значения и порядок величин важнейших параметров материалов.

Каждый студент выполняет эти задания в соответствии с индивидуальным вариантом исходных данных, который определяется преподавателем на установочной консультации, либо по номеру студента в списке группы.

Правильность решения оценивается преподавателем на практических занятиях, поэтому ответы на задачи в учебном пособии не приведены.

Задания рекомендуется оформлять в отдельной тетради либо на листах формата А4 (210×297). При проведении расчетов следует приводить общие и окончательные расчетные формулы, выполнять в явном виде подстановку числовых данных, выраженных обязательно в одной системе единиц. При решении задач по физическим основам микро- и наноэлектроники допускается не соблюдать СИ, т. к. размерность основных параметров полупроводников в большинстве справочников указана в сантиметрах. Числовой ответ обязательно должен иметь размерность (наименование единиц измерения). Большую часть задач достаточно решать с точностью до 2–3 знаков после запятой.

Пример оформления задачи

Через пластину германия с удельным сопротивлением $0,01 \text{ Ом}\cdot\text{м}$ проходит электрический ток плотностью 10 мА/мм^2 . Найти среднюю скорость дрейфа электронов и дырок, если их подвижности $3800 \text{ см}^2/(\text{В}\cdot\text{с})$

и $1800 \text{ см}^2/(\text{В} \cdot \text{с})$ соответственно. (Обратите внимание на размерность числовых данных!)

Плотность дрейфового тока определяется по закону Ома $j_{\text{др}} = \sigma E = \sigma / \rho$, средняя скорость дрейфа — $V_{\text{др}} = \mu E$. Средняя скорость дрейфа электронов — $V_{\text{др}n} = \mu_n E = \mu_n j_{\text{др}} \rho$, средняя скорость дрейфа дырок — $V_{\text{др}p} = \mu_p E = \mu_p j_{\text{др}} \rho$.

Преобразуем размерность числовых данных в единую систему, например, в СИ:

$$j_{\text{др}} = 10 \text{ мА} / \text{мм}^2 = 10 \cdot 10^{-3} / 10^{-6} \text{ А/м} = 10^4 \text{ А/м},$$

$$\mu_n = 3800 \text{ см}^2/(\text{В} \cdot \text{с}) = 3800 \cdot 10^{-4} = 0,38 \text{ м}^2/(\text{В} \cdot \text{с}),$$

$$\mu_p = 1800 \text{ см}^2/(\text{В} \cdot \text{с}) = 1800 \cdot 10^{-4} = 0,18 \text{ м}^2/(\text{В} \cdot \text{с}).$$

$$V_{\text{др}n} = 0,38 \cdot 10^{-4} \cdot 0,01 = 38 \text{ м/с}, \quad V_{\text{др}p} = 0,18 \cdot 10^{-4} \cdot 0,01 = 18 \text{ м/с}.$$

Ответ: средняя скорость дрейфа электронов 38 м/с, дырок 18 м/с.

Однородный полупроводник

Задача 1

Исходные данные к задаче 1

Вариант	Тип примеси	$N_{\text{пр}1}, \text{см}^{-3}$	$N_{\text{пр}2}, \text{см}^{-3}$
1	Донорная	10^{16}	10^{18}
2	Акцепторная	10^{15}	10^{17}
3	Донорная	$5 \cdot 10^{18}$	$5 \cdot 10^{15}$
4	Акцепторная	10^{18}	10^{16}
5	Донорная	10^{17}	10^{15}
6	Акцепторная	$5 \cdot 10^{16}$	$5 \cdot 10^{17}$
7	Донорная	$5 \cdot 10^{16}$	$5 \cdot 10^{17}$
8	Акцепторная	10^{15}	10^{18}
9	Донорная	10^{15}	10^{17}
10	Акцепторная	$5 \cdot 10^{16}$	$5 \cdot 10^{18}$
11	Донорная	$5 \cdot 10^{18}$	$5 \cdot 10^{16}$
12	Акцепторная	$5 \cdot 10^{17}$	$5 \cdot 10^{15}$
13	Донорная	10^{16}	10^{17}
14	Акцепторная	10^{16}	10^{17}
15	Донорная	10^{18}	10^{15}

Определите концентрацию основных и неосновных носителей заряда при комнатной температуре ($T = 300 \text{ K}$) в образце германия для двух заданных значений концентраций примеси.

Определите концентрацию основных и неосновных носителей заряда при комнатной температуре ($T = 300 \text{ K}$) в образце кремния для тех же значений концентраций примеси.

Сравните и проанализируйте полученные результаты.

Задача 2

Определите подвижность дырок в заданном полупроводнике при заданной температуре.

Исходные данные к задаче 2

Вариант	Материал полупроводника	$T, ^\circ\text{C}$
1	Германий	50
2	Кремний	60
3	Германий	70
4	Кремний	80
5	Германий	40
6	Кремний	50
7	Германий	60
8	Кремний	70
9	Германий	80
10	Кремний	40
11	Германий	10
12	Кремний	-10
13	Германий	-20
14	Кремний	-30
15	Германий	10

Задача 3

Определите проводимость и положение уровня Ферми для заданного типа полупроводника, легированного заданной примесью ($T = 300 \text{ K}$).

Исходные данные к задаче 3

Вариант	Материал полупроводника	Тип примеси	$N_{\text{пр}}, \text{см}^{-3}$
1	Германий	Донорная	10^{16}
2	Германий	Акцепторная	10^{15}
3	Кремний	Донорная	$5 \cdot 10^{17}$
4	Кремний	Акцепторная	$5 \cdot 10^{18}$
5	Германий	Донорная	10^{17}
6	Германий	Акцепторная	10^{18}
7	Кремний	Донорная	10^{16}
8	Кремний	Акцепторная	10^{15}
9	Германий	Донорная	$5 \cdot 10^{14}$
10	Германий	Акцепторная	$5 \cdot 10^{14}$
11	Кремний	Донорная	10^{17}
12	Кремний	Акцепторная	10^{18}
13	Германий	Донорная	$5 \cdot 10^{16}$
14	Германий	Акцепторная	$5 \cdot 10^{15}$
15	Кремний	Донорная	10^{18}

Задача 4

Определите количество дырок в заданном объеме собственного полупроводника при заданной температуре.

Сколько электронов находится в заданном полупроводнике?

Исходные данные к задаче 4

Вариант	Материал полупроводника	Объем, мкм^3	$T, ^\circ\text{C}$
1	Германий	30	50
2	Кремний	300	60
3	Германий	50	70
4	Кремний	100	80
5	Германий	40	40
6	Кремний	400	50
7	Германий	30	60
8	Кремний	200	70
9	Германий	60	80

Вариант	Материал полупроводника	Объем, мкм ³	T, °C
10	Кремний	500	40
11	Германий	250	50
12	Кремний	350	60
13	Германий	50	70
14	Кремний	150	80
15	Германий	450	50

Задача 5

Какое напряжение нужно приложить к стержню примесного полупроводника длиной 5 см, чтобы его основной носитель заряда дрейфовал вдоль стержня с заданной скоростью при комнатной температуре? (Материал полупроводника и название примеси заданы.)

Исходные данные к задаче 5

Вариант	Материал полупроводника	Название примеси	Скорость дрейфа, м/с
1	Германий	Фосфор	150
2	Германий	Алюминий	140
3	Кремний	Фосфор	170
4	Кремний	Алюминий	160
5	Германий	Сурьма	180
6	Германий	Бор	160
7	Кремний	Сурьма	150
8	Кремний	Бор	180
9	Германий	Индий	150
10	Германий	Мышьяк	190
11	Кремний	Индий	130
12	Кремний	Мышьяк	200
13	Германий	Галлий	140
14	Германий	Сурьма	170
15	Кремний	Галлий	150

Задача 6

В кристалле полупроводника при заданной температуре длина диффузии носителя заряда $L = 2$ мм. Определить подвижность и время жизни носителей заряда.

Исходные данные к задаче 6

Вариант	Материал полупроводника	Тип носителя	T , К
1	Кремний	Электрон	200
2	Германий	Электрон	200
3	Кремний	Электрон	250
4	Германий	Электрон	250
5	Кремний	Дырка	200
6	Германий	Дырка	200
7	Кремний	Дырка	250
8	Германий	Дырка	250
9	Кремний	Электрон	350
10	Германий	Электрон	350
11	Кремний	Дырка	350
12	Германий	Дырка	350
13	Кремний	Электрон	300
14	Германий	Электрон	300
15	Кремний	Дырка	350

Задача 7

В однородный полубесконечный полупроводник p -типа в плоскости $x = 0$ непрерывно вводится и поддерживается постоянной во времени избыточная концентрация электронов $\Delta n(x = 0)$. Определите плотность электронного тока при $x = 0$ ($T = 300$ К). Следует считать задачу одномерной.

Исходные данные к задаче 7

Вариант	Материал полупроводника	Время жизни, с	Δn , см ⁻³
1	Кремний	10^{-3}	10^{15}
2	Германий	10^{-3}	10^{15}
3	Кремний	10^{-4}	10^{15}

Вариант	Материал полупроводника	Время жизни, с	Δn , см ⁻³
4	Германий	10^{-4}	10^{15}
5	Кремний	10^{-3}	10^{14}
6	Германий	10^{-3}	10^{14}
7	Кремний	10^{-4}	10^{14}
8	Германий	10^{-4}	10^{14}
9	Кремний	10^{-3}	10^{13}
10	Германий	10^{-3}	10^{13}
11	Кремний	10^{-4}	10^{13}
12	Германий	10^{-4}	10^{13}
13	Кремний	10^{-3}	10^{16}
14	Германий	10^{-3}	10^{16}
15	Кремний	10^{-4}	10^{16}

Задача 8

Определите ток через образец полупроводника прямоугольной формы размерами $l \times b \times h = 5 \times 2 \times 1$ мм, если вдоль образца приложено напряжение U . Известна концентрация свободных носителей заряда n и их подвижность μ_n .

Исходные данные к задаче 8

Вариант	n , см ⁻³	μ_n , см ² /(В·с)	U , В
1	10^{15}	400	50
2	10^{15}	500	40
3	10^{15}	1000	30
4	10^{15}	700	50
5	10^{16}	400	40
6	10^{16}	500	30
7	10^{16}	900	50
8	10^{16}	700	40
9	10^{16}	800	30
10	10^{16}	600	60
11	10^{17}	400	10
12	10^{17}	500	20
13	10^{17}	600	30

Вариант	$n, \text{см}^{-3}$	$\mu_n, \text{см}^2/(\text{В}\cdot\text{с})$	$U, \text{В}$
14	10^{17}	700	40
15	10^{17}	800	50

Задача 9

Вычислите отношение дырочной составляющей тока через полупроводник к полному току ($T = 300 \text{ К}$):

- в собственном полупроводнике;
- в полупроводнике p -типа с заданным удельным сопротивлением ρ .

Исходные данные к задаче 9

Вариант	Материал	$\rho, \text{Ом}\cdot\text{м}$
1	Кремний	0,05
2	Германий	0,05
3	Кремний	0,04
4	Германий	0,04
5	Кремний	0,06
6	Германий	0,06
7	Кремний	0,03
8	Германий	0,03
9	Кремний	0,02
10	Германий	0,02
11	Кремний	0,01
12	Германий	0,01
13	Кремний	0,07
14	Германий	0,07
15	Кремний	0,04

Электронно-дырочный переход

Задача 1

В результате измерений было установлено, что при комнатной температуре ток германиевого диода при $U = -1 \text{ В}$ составил $I = 100 \text{ мкА}$. Определите ток при напряжениях U_1, U_2 при комнатной температуре. (Током утечки и влиянием базы следует пренебречь.)

Исходные данные к задаче 1

Вариант	$U_1, \text{В}$	$U_2, \text{В}$
1	-0,2	+0,2
2	-0,3	+0,3
3	-0,4	+0,4
4	-0,2	+0,3
5	-0,3	+0,4
6	-0,5	+0,5
7	-0,2	+0,4
8	-0,5	+0,3
9	-0,4	+0,5
10	-0,3	+0,2
11	-0,6	+0,3
12	-0,6	+0,4
13	-0,7	+0,3
14	-0,7	+0,4
15	-0,8	+0,5

Задача 2

Обратный ток идеального диода $I_{\text{обр}} = 5 \cdot 10^{-6} \text{ А}$. Какое напряжение нужно приложить к диоду, чтобы ток диода составил величину I_1 ?

Исходные данные к задаче 2

Вариант	$I_1, \text{мА}$	Вариант	$I_1, \text{мА}$
1	75	9	85
2	50	10	95
3	100	11	40
4	90	12	30
5	80	13	105
6	70	14	110
7	60	15	120
8	65		

Задача 3

Определите сопротивление базы германиевого диода, если известна величина прямого тока I_1 и соответствующее ему напряжение U_1 и величина тока насыщения I_0 .

Исходные данные к задаче 3

Вариант	I_1 , мА	U_1 , В	I_0 , мкА
1	100	0,58	25
2	100	0,65	25
3	100	0,58	20
4	70	0,50	25
5	70	0,48	25
6	70	0,50	20
7	100	0,65	20
8	70	0,48	20
9	50	0,45	25
10	50	0,45	20
11	50	0,65	30
12	80	0,65	30
13	80	0,55	30
14	80	0,60	25
15	60	0,65	25

Задача 4

В кремниевом резком p - n переходе с концентрациями примесей N_a и N_d лавинный пробой наступает при напряженности электрического поля E_1 . Вычислите ширину p - n перехода и обратное напряжение, необходимое для начала возникновения пробоя.

Исходные данные к задаче 4

Вариант	N_a , м^{-3}	N_d , м^{-3}	E_1 , В/м
1	10^{20}	10^{22}	$6 \cdot 10^7$
2	10^{20}	10^{22}	$7 \cdot 10^7$
3	10^{20}	10^{22}	$6,5 \cdot 10^7$
4	10^{20}	10^{22}	$7,5 \cdot 10^7$
5	10^{19}	10^{21}	$7,5 \cdot 10^7$

Вариант	$N_a, \text{м}^{-3}$	$N_d, \text{м}^{-3}$	$E_l, \text{В/м}$
6	10^{20}	10^{22}	$5,5 \cdot 10^7$
7	10^{20}	10^{22}	$5 \cdot 10^7$
8	10^{19}	10^{21}	$8 \cdot 10^7$
9	10^{19}	10^{21}	$8,5 \cdot 10^7$
10	10^{19}	10^{21}	$7 \cdot 10^7$
11	10^{20}	10^{21}	$7,5 \cdot 10^7$
12	10^{20}	10^{22}	$6,5 \cdot 10^7$
13	10^{21}	10^{19}	$7 \cdot 10^7$
14	10^{21}	10^{19}	$6,5 \cdot 10^7$
15	10^{22}	10^{20}	$7,5 \cdot 10^7$

ОГЛАВЛЕНИЕ

Предисловие	3
1. Физический эффект и его компоненты	4
1.1. Определение физического эффекта	4
1.2. Физические эффекты в электронике	6
2. Однородный полупроводник	8
2.1. Собственный полупроводник	8
2.2. Электронный полупроводник	24
2.3. Дырочный полупроводник	33
3. Уровень Ферми.....	40
Задачи к практическому занятию № 1	52
4. Дрейф. Электропроводность.....	54
5. Генерация и рекомбинация в полупроводниках.....	65
6. Диффузионный ток. Законы движения носителей заряда в полупроводниках.....	76
Задачи к практическому занятию № 2	83
7. Применение однородных полупроводников	85
7.1. Варисторы.....	86
7.2. Терморезисторы	99
7.3. Фоторезисторы.....	113
8. p - n Переход	129
8.1. Определение и классификация p - n переходов	129
8.2. Технология производства p - n перехода.....	131
8.3. Равновесное состояние p - n перехода.....	143
8.4. Ток в p - n переходе в равновесном состоянии	149

8.5. Контактная разность потенциалов.....	151
8.6. Энергетическая диаграмма p - n перехода в равновесном состоянии	153
8.7. Неравновесное состояние p - n перехода	156
8.8. Идеальный p - n переход	165
8.9. Реальный p - n переход.....	177
8.10. Эквивалентные схемы реального p - n перехода	187
8.11. Пробой p - n перехода	189
8.12. Туннелирование в сильнолегированных p - n переходах.....	200
Задачи к практическому занятию № 3	216
Задачи к практическому занятию № 4	217
9. Гетеропереходы	219
9.1. Общие сведения о гетеропереходах.....	219
9.2. Энергетическая диаграмма гетероперехода	220
9.3. Прямосмещенный гетеропереход.....	223
9.4. Получение гетеропереходов.....	224
9.5. Электрические свойства гетеропереходов	225
9.6. Применение гетеропереходов в оптоэлектронике.....	227
Библиографический список	231
Приложение. Домашние задания	233

Учебное издание

Дурнаков Андрей Адольфович

ФИЗИЧЕСКИЕ ОСНОВЫ МИКРО- И НАНОЭЛЕКТРОНИКИ

Редактор И. В. Меркурьева
Компьютерный набор А. А. Дурнакова
Верстка Е. В. Ровнушкиной

Подписано в печать 20.02.2020. Формат 70×100 1/16.
Бумага писчая. Цифровая печать. Усл. печ. л. 20,0.
Уч.-изд. л. 13,5. Тираж 30 экз. Заказ 24.

Издательство Уральского университета
Редакционно-издательский отдел ИПЦ УрФУ
620049, Екатеринбург, ул. С. Ковалевской, 5
Тел.: 8 (343) 375-48-25, 375-46-85, 374-19-41
E-mail: rio@urfu.ru

Отпечатано в Издательско-полиграфическом центре УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: 8 (343) 358-93-06, 350-58-20, 350-90-13
Факс: 8 (343) 358-93-06
<http://print.urfu.ru>

Для заметок

